

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2005, 08:31:17 ; Search time 84 Seconds
(without alignments)
1944.682 Million cell updates/sec

Title: US-10-606-300-11

Perfect score: 1619

Sequence: 1 MTKIELRAGNLTGLKVSAGV.....VEAILEPVKNLTWPSGIHQN 319

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1619	100.0	319	2	Q81884 arabidopsis
2	1348	83.3	319	2	Q84L11 actinidia d
3	1322	81.7	324	2	Q84L20 malus domes
4	1261	77.9	322	2	Q6bdj2 spinacia ol
5	695	42.9	310	2	Q8a1P5 bacteroides
6	625	38.6	347	2	Q7UZ40 rhodospirillum rubrum
7	593.5	36.7	345	2	Q9VGF3 drosophila
8	593	36.6	342	2	Q7QHC3 anopheles g
9	544.5	33.6	342	2	Q9VGF1 drosophila
10	503	31.1	439	2	Q20127 caenorhabditis elegans
11	426	26.3	336	2	Q8PF04 xanthomonas
12	420.5	26.0	294	2	Q9VGF2 drosophila
13	417.5	25.8	335	2	Q8P3K6 xanthomonas
14	414.5	25.6	336	2	Q6FU59 candida gla
15	413	25.5	330	2	Q6BZU5 yarrowia li
16	387.5	23.9	359	2	Q75DD3 aebhya goss
17	383	23.7	328	2	Q82N99 streptomyces
18	378	23.3	433	2	Q7B6W9 neurospora
19	375	23.2	328	2	Q6CIG5 kluyveromyces fragilis
20	369.5	22.8	328	2	Q826A9 streptomyces
21	366	22.6	329	2	Q52472 pseudomonas
22	362.5	22.4	297	2	Q9KE47 bacillus ha
23	361	22.3	325	2	Q89VC3 bradyrhizobium
24	358	22.1	335	1	Q84212 saccharomyces cerevisiae
25	349	21.6	306	2	Q65HS3 bacillus li
26	346.5	21.4	329	2	Q8UA47 agrobacterium tumefaciens
27	345.5	21.4	434	2	Q7XEJ9 oryza sativ
28	345	21.3	369	2	Q6BPT1 debaryomyces hansenii
29	344.5	21.3	334	2	Q9RJW1 streptomyces
30	342.5	21.2	326	2	Q7NMC8 gloeobacterium
31	342.5	21.2	329	2	Q8ETF4 oceanobacillus

32	339	20.9	304	2	Q81820 bacillus ce
33	337.5	20.8	306	1	YQKF_BACSU
34	336.5	20.8	314	2	Q7NKG3 gloeobacter
35	329.5	20.4	304	2	Q81MD1 bacillus an
36	329.5	20.4	304	2	Q6HE69 bacillus th
37	327.5	20.2	352	2	Q6F726 acinetobact
38	327	20.2	352	2	Q6DAK3 erwinia car
39	325.5	20.1	304	2	Q731K1 bacillus ce
40	323.5	20.0	304	2	Q63515 bacillus ce
41	323.5	20.0	339	2	Q92TD8 rhizobium m
42	317	19.6	333	2	Q64UL3 bacteroides
43	317	19.6	351	2	Q8UAD6 asaccharobacter
44	315	19.5	326	2	Q8FGX5 escherichia
45	314	19.4	325	2	Q8G5A3 bifidobacte

ALIGNMENTS

RESULT 1

Q81884 PRELIMINARY; PRT; 319 AA.
AC Q81884;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE Hypothetical protein T16L1.160 (AT4G33670/T16L1_160) (L-galactose dehydrogenase) (Hypothetical protein AT4G33670).
GN Name=T16L1.160; Synonyms=AT4G33670, L-GalDH;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Obermaier B., Deutschenbaur S., Piravandi E., Hoheisel J., Jesse T., Heijnen L., Vos P., Mewes H.W., Mayer K.F.X., Schueller C.; Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Koesena E., Chen H., Cheuk R., Kim C.J., Meyers M.C., Shinn P., Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A., Ecker J.R.; Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J., Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S., Ecker J.R.; Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=22043773; PubMed=12047629;
RT "Antisense suppression of l-galactose dehydrogenase in Arabidopsis thaliana provides evidence for its role in ascorbate synthesis and reveals light modulated l-galactose synthesis.";
RT Plant J. 30:541-553(2002).

Matches	248;	Conservative	40;	Mismatches	31;	Indels	0;	Gaps	0;		
Qy	1	MTKIELRALNGTGLKVS	SAVGFS	PLGVS	VFGPVA	EDDAV	TVREA	FLRG	INFFDTSPY	60	
Db	6	LPKVELRELNGTGLKLS	CVGFS	PLGNV	FGPVS	DDAIG	SVREA	PRRG	INFFDTSPY	65	
Qy	61	GTLSEKWLKGLKALQV	PRSDY	IVATK	CGRYK	EGDF	SAEVR	KSID	ESLERLQDYVDI	120	
Db	66	GTLSEKVLGKTLKALG	VPRSEY	VVATK	CGRYA	DGDF	SADRV	TKSID	ESLERLQDYVDI	125	
Qy	121	LHCHDIEFGSLDQIV	SETTIP	ALQKL	KQEGK	TRFIG	ITGLP	LIDIT	FYVLDR	VPGVVDVIL	180
Db	126	LQCHDIEFGSLDQIV	NETIIP	ALKKL	KEAGK	IRFIG	ITGLP	LUGIT	FYVLDR	IPAGMVDVVL	185
Qy	181	SYCHYGNDSTLLDL	LPYLKS	KGVGVI	SASPL	AMGLLT	EQQPP	HPWH	PASPELKS	ASKAAV	240
Db	186	SYCHYINDSTLEDLI	LPYLKS	KGVIIS	ASPLA	MGLLT	TENG	PPWH	PASAE	LKSACRAA	245
Qy	241	AHCKSGKKITTKLAL	QVLSANKEI	ISSVL	GVMSV	SQVSE	ENVA	AVTE	LESIG	MDQETISEV	300
Db	246	VYCKERGNNISKLAL	QVLSLNKDI	ISSVL	GVNWS	INQVE	ENVA	AAV	ELATIG	KEKILAEV	305
Qy	301	EAILPVPKNLTWPSG	IGHQN	319							
Db	306	EAILKPVKNOTWPSG	LQOOS	324							

RESULT 4						
Q6BDJ2	ID	Q6BDJ2	PRELIMINARY;	PRT;	322 AA.	
AC	Q6BDJ2;					
DT	25-OCT-2004	(T-EMBLrel. 28,	Created)			
DT	25-OCT-2004	(T-EMBLrel. 28,	Last sequence update)			
DT	25-OCT-2004	(T-EMBLrel. 28,	Last annotation update)			
DE	L-galactose dehydrogenase.					
GN	Name=GDH;					
OS	Spinacia oleracea (Spinach).					
OC	Eukaryota; Viridiplantae;	Streptophyta; Embryophyta;	Tracheophyta;			
OC	Spermatophyta; Magnoliophyta;	eudicotyledons; core eudicots;				
OC	Caryophyllales; Amaranthaceae;	Spinacia.				
OX	NCBI_TaxID=3562;					
ON	[1]					
RP	SEQUENCE FROM N.A.					
RA	Mieda T., Yabuta Y., Rapolu M., Motoki T., Ishikawa T., Yoshimura K., Shigeoka S.;					
RA	"Direct Submission.";					
RL	Submitted (JAN-2004) to the EMBL/GenBank/DBSJ databases.					
DR	EMBL; AB160990; BAD32687.1; -.					
DR	InterPro; IPR001395; Aldo/ket red.					
DR	Pfam; PF00248; Aldo ket red; 1.					
DR	PRINTS; PR00069; ALDKETREDTASE.					
DR	PRODOM; PD000286; Aldo/ket red; 1.					
SQ	SEQUENCE 322 AA; 35262 MW; CB5BA18D81A27AF4 CRC64;					
Query Match		77.9%;	Score 1261;	DB 2;	Length 322;	
Best Local Similarity		75.0%;	Pred. No. 5.7e-82;			
Matches 237;		Conservative	38;	Mismatches 41;	Indels 0;	Gaps 0;
Qy	3	KIEALRGNTGLKVSAVGFGASPLGSGVFPVAEDDAVATVREAFLRGINFFDTSPYYGGT	62			
Dd	6	KLERRELNGTGLNLSCVGFGASPLGNGVFDVSEEQSIATVTEAFNQGINFFDTSPYYGAT	65			
Qy	63	LSEKWLKGKIALGVPPRSDIYIVATKCGRYKEGDFSAERVKSIDESLERLOLDVDILH	122			
Dd	66	LSEKVLGKCIKALGASRDYIVATKCGRYAGDFSAERVTKSIDESLERLQLEVEVDILQ	125			
Qy	123	CHDIEFGSLSDOIVSETIPALQKLQEGKTREIGITGLPLDIFTVLDRVPVPGTVDDVLVS	182			
Dd	126	CHDIEFGSLSDOIVNETIPALQKIKESGKTREIGITGLPLEVYTYVLDVPEGTIDVVLVS	185			
Qy	183	CHYGVDNSTDLDDLPLYLKSGVGVISAPSLANGLLTBOGPPEWHIPASPELKSAAVAH	242			
Dd	186	CHYCINDSTLEDMLPYFKSKGVGVINASPISMGHLHTNGPPEWHIPASPEIKAACKAAADY	245			

RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

RN [6]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003694; AAF54729.1; -;
DR HSP; F52895; IJ96.
DR FlyBase; FBgn0037973; CG18547.
DR InterPro; IPR001395; Aldo/ket red.
DR Pfam; PF00248; Aldo.ket.red; I.
DR PRINTS; PR00069; ALDKETDASE.
DR PRODOM; PD000288; Aldo/ket_red; 1.
SQ SEQUENCE 345 AA; 38868 MW; 86124CBB34BF04F2 CRC64;

Query Match 36.7%; Score 593.5; DB 2; Length 345;
Best Local Similarity 42.5%; Pred. No. 3.1e-34;
Matches 128; Conservative 63; Mismatches 99; Indels 11; Gaps 7;

```
QY 3 KIELRALGNTGLKVSAGVFGASPLGSVFGPVAEDDAVATVREAFRLGINFDTSPYYGGT 62
DB 21 RMEYENLAKTGLQVSKVSGFGGALCANYGFDL-EGIKTVHEAVKSGINYIDTAPWYGG 79
QY 63 LSEKMLGKGLKALQVPRSDIYVATKCGRYK----EGDFSAERVKSDIESLERLQLDYV 118
DB 80 RSEEVIGLALK--DVPRSYIATKVARYELDYDQDFSAKTKRESVERSLKLLGLDYV 137
QY 119 DILHCHDIEFG-SLDQIVSETIPALQKLQKQKTRFIGITGLPIDIFTYVLDVPPPGTVD 177
DB 138 DVIQIHDIHFAKDLDIVINETLPTLEQLVKEGKARFIGVSAYPSVLKEFLTRT-AGRLD 196
QY 178 VILSYCHYGVDNSTLLDLLPYLKSQGVIGASPLAMGLLTQGGPPEWHPASPELKSASK 237
DB 197 TVLTAVRTLTDETLLEVDLPKQNLGVCIAAAHALGLLTNAGPQWHPASDEQKALAR 256
QY 238 AAVAHCKSKGKITKALQYSLAN-KEISSVLVGNSSVSQVEENVAATVTELSLGMDOET 296
DB 257 KASEVCKERGVELGKLAITYTMSGLPEVSTFLTGMQTRQLLRINLDA-NEVGLSDKEQEV 315
QY 297 L 297
DB 316 L 316
```

RESULT 8

```
Q7QHC3 PRELIMINARY; PRT; 342 AA.
AC Q7QHC3;
DT 01-MAR-2004 (TREMRELrel. 26, Created)
DT 01-MAR-2004 (TREMRELrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMRELrel. 26, Last annotation update)
DE Eb1P8370 (Fragment).
GN Name=ebiG8370; ORFNames=ENSANGSG0000006316;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AA80100816; EAA05276.1; -.
DR HSP; F52895; IJ96.
DR InterPro; IPR001395; Aldo/ket red.
DR Pfam; PF00248; Aldo.ket.red; I.
DR PRINTS; PR00069; ALDKETDASE.
DR PRODOM; PD000288; Aldo/ket_red; 1.
FT NON_TER 342
FT NON_TER 342
SQ SEQUENCE 342 AA; 38590 MW; 52B38ADD54170598 CRC64;
```

```
Query Match 36.6%; Score 593; DB 2; Length 342;
Best Local Similarity 39.7%; Pred. No. 3.4e-34;
Matches 126; Conservative 68; Mismatches 105; Indels 18; Gaps 7;
QY 3 KIELRALGNTGLKVSAGVFGASPLGSVFGPVAEDDAVATVREAFRLGINFDTSPYYGGT 62
DB 14 RMEYENLAKTGLKVSQVSGFGPFSQLYGNLDEAKALEAVRHAVQRGINYIDTAPYGG 73
QY 63 LSEKMLGKGLKALQVPRSDIYVATKCGRYK---EG-PDFAERVKSDIESLERLQLDYV 118
DB 74 RSEEVIGLALK--IPRAAYVATKVSYERYEDGDFYSQKTRQSVRSVLSLGLDYL 131
QY 119 DILHCHDIEFG-SLDQIVSETIPALQKLQKQKTRFIGITGLPIDIFTYVLDVPPPGTVD 177
DB 132 DVQIHDIHFAKDLDIVINETLPTLEALRASGKLRFIGVSAVPLAVLREIITKA-PGRFD 190
QY 178 VILSYCHYGVDNSTLLDLLPYLKSQGVIGASPLAMGLLTQGGPPEWHPASPELKSASK 237
DB 191 TVLSYCRNTLPDDSLBEYIPFPEHKGMLICASGHGMLLTNGGPPQWHPADRLRRECA 250
QY 238 AAVAHCKSKGKITKALQYSLANKEISSVLVGNSSVSQVEENVAATVTELSLGMDOETL 297
DB 251 EAAAYCKREGVELGKLAHMLQMPGATFLAGMTPELVQINLDAYF-----EGL 301
QY 298 SEVEA-ILEPVKNLTWP 313
DB 302 SEKEADVLSYKERVFP 318
```

RESULT 9

```
Q9VGF1 PRELIMINARY; PRT; 342 AA.
ID Q9VGF1
AC Q9VGF1;
DT 01-MAY-2000 (TREMRELrel. 13, Created)
DT 01-MAY-2000 (TREMRELrel. 13, Last sequence update)
DT 25-OCT-2004 (TREMRELrel. 28, Last annotation update)
DE CG3397-PA (GH17980P).
GN ORFNames=CG3397;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocke J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gaber G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
```

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Paclebb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskaas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the *Drosophila*
melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskaas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Paclebb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003694; AAF54731.1; -;
DR EMBL; AV118797; AAM50657.1; -;
DR IntAct; Q9VGF1; -;
DR FlyBase; Fgn0037975; CG3397.
DR InterPro; IPR001395; Aldo/ket_red.
DR Pfam; PF00248; Aldo_ket_red; 1.
DR PRINTS; PR00069; ALDKETRDASE.
SQ SEQUENCE 342 AA; 38085 MW; D9128A46172B76B1 CRC64;

Query Match 33.6%; Score 544.5; DB 2; Length 342;
Best Local Similarity 36.6%; Pred. No. 9.8e-31;
Matches 117; Conservative 71; Mismatches 117; Indels 15; Gaps 7;
QY 3 KIELRALQNTGLKYSVAGFSGPLGVSFG-PAVEDAVATVREAFRLGINFDTSPYYGG 61
DB 21 RMEYRQJGSLRYSKALGATLSKLFSDDFDEEGILTVQEAIRSGINVIDTAPFYGG 80
QY 62 TLSEKMLGKGLKALQVPRSDYIVATKGRYK-----EGPDSAEVRKSIDSLERLQLDY 117
DB 81 GKSEBELLQALK--DVPREAYIATKVARYELDPNNMEDYTAAKARESIVKRSLELLQLDR 138
QY 118 VDILHCHDIEFG-SLDQIVSETIPALQKLQEGKTRFIGITGLPDIITYVLDRVPPGTV 176
DB 139 VDVLQVHDVDAAPSLDMVLNETIPVLEBYVQAGKARFIGVTAYDDVVLKECAER-GKGRI 197
QY 177 DVILSYCHYGVNDSTLIDLPLKSKGVVISASPLAMGLLTEQPPPEWHKSPKLSAS 236
DB 198 QVVLNARYTLLDNTLLRHKAFQEMGVGVVCAAAHSIGLLSNAGPQSHFGSPPELLAVG 257
QY 237 KAAVAHCKSKGKTKALQYSLANKEISSVLGVSSVSQVEENVAATELESIGMDQET 296
DB 258 KRGAEICQKRVNVELGKLAAVYTMQDGAATELIGIPNKKLRINLDAIFD---GLTSHE 313
QY 297 LSEVEATILEPV--KNLTWPS 314
DB 314 QEVLYLRENVFTKSYSGWS 333
RESULT 10
Q20127 PRELIMINARY; PRT; 439 AA.
ID Q20127
AC Q20127
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein F37C12.12.
GN ORFNames=F37C12.12;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RG WormBase Consortium;
RT "Genome sequence of the nematode *C. elegans*: a platform for
investigating biology. The *C. elegans* Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Fulton L.;
RT "The sequence of *C. elegans* cosmid F37C12.";
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; U00033; AAC48300.1; -;
DR PIR; T28841; T28841.
DR WormBase; WBGene000031176; F37C12.12.
DR WormPeP; F37C12.12; CE00825.
DR InterPro; IPR001395; Aldo/ket_red.
DR Pfam; PF00248; Aldo_ket_red; 1.
DR PRINTS; PR00069; ALDKETRDASE.

```

DR ProDom; PD000288; Aldo/ket_red; 1.
KW Hypochemical protein.
SQ SEQUENCE 439 AA; 49696 MW; E0508EE84720FFCD CRC64;

Query Match          31.1%; Score 503; DB 2; Length 439;
Best Local Similarity 36.8%; Pred. No. 1.2e-27;
Matches 120; Conservative 58; Mismatches 106; Indels 42; Gaps 10;

QY      3 KIELRALGNTGLKVS AVFGASPLGS VFGPAEDDAVATREAFRLGINFDPSPYYGGT 62
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      82 RMNYRQIQPTDIRMSKIGFCGAAATGGMFGNV -EDSIKIIVETAIKQGINTIDGYWYSQS 140
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     63 LSEKMLKGKIKALQVRPSDIVATKGR-----YKEGFDSAERVKRKIDSLESLQLDYV 118
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     141 RSEIIIGKALS--KIPKAYIISTKVGRFELDVRTDFRADKILLESITNSLKKLQITYI 198
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY    119 DI--LHCHDIEFGSLDIQIV-SETIPALQKLQEGEKTRFIGITGLPLD----IFTYVLDRV 171
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    199 DICVQIHADAPNBSIVLYETLQALEMAKSGKIRHIGLTGYPLKLVYSPRYVN-- 256
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY    172 PPGTVDVILSYCHGVNDSTLLDLLPYLKSGGVVISASPAMGLLTTEQQGPPEWHPASPE 231
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    257 -----APKLYFSVNLST-----TRNTAVINSGALCWGLLTKEKGGPPMHHPASDE 299
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY    232 LKSASKAAVAHCCKSGKKITKLALQYSLANKETSISVLVGSSYSQVSENV-----AAVTE 286
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    300 IKAECIAAATTCYCSKNISISKLAUDYALNPVNICCLVGVGDSVQQVLDNLNELSFRITD 359
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY    287 LESLGMDQETLSVEAILEPVKNLTW 312
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    360 VEORVRDR-----IMRRYLDRLENAGW 381
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT	ID	Q8PF04	PRELIMINARY;	PRT;	336 AA.
AC	Q8PF04;				
DT	01-OCT-2002	(T-EMBLrel. 22, Created)			
DT	01-OCT-2002	(T-EMBLrel. 22, Last sequence update)			
DT	01-OCT-2002	(T-EMBLrel. 22, Last annotation update)			
DE	Oxidoreductase.				
GN	OrderedLocusNames=XAC4184;				
OS	Xanthomonas axonopodis (pv. citri).				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;				
OC	Xanthomonadaceae; Xanthomonas.				
OX	NCBI_TaxID=92829;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=306 / ATCC 13902 / XV 101;				
RC	MEDLINE=220221145; PubMed=12042127; DOI=10.1038/417459a;				
RX	da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,				
RA	Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,				
RA	Alves L.M.C., do Amaral A.M., Bartolini M.C., Camargo L.E.A.,				
RA	Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,				
RA	Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,				
RA	Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,				
RA	Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,				
RA	Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,				
RA	Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,				
RA	Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,				
RA	Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,				
RA	Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,				
RA	Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,				
RA	Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,				
RA	Stebul J.C., Kitajima J.P.;				
RT	"Comparison of the genomes of two Xanthomonas pathogens with differing				
RT	host specificities."				
RL	Nature 417:459-463 (2002).				
RL	EMBL; AE012069; AAM39019.1; -.				
DR	HSSP; P52895; IJ96.				
DR	Interpro; IPR001395; Aldo/ket_red.				
DR	PFam; PF00248; Aldo_ket_red; 1.				
DR	ProDom; PD000288; Aldo/ket_red; 1.				

KW	Complete proteome.
SQ	SEQUENCE 336 AA; 34982 MW; 8BF6EA56C8E6BE97 CRC64;
	Query Match 26.3%; Score 426; DB 2; Length 336;
	Best Local Similarity 33.7%; Pred. No. 2,8e-22;
	Matches 118; Conservative 60; Mismatches 104; Indels 68; Gaps 12;
QY	1 MYKIELRALGNTGLKVSAGVFGASPLGSVFGVAEDDAVATREAPRLGINPDTSPYYG 60
Dd	1 MSGVARQLGGTDVOLSLGFGAAPTGINLYTEVDEADALAAVAGAFAEGIRHFDTPAPPY 60
QY	61 GTLSEKMLCKGLKALQVPDSDIYATKCGRY-----KEG-----DFS AER 101
Dd	61 YGLSEARLGRGLAG--VPRADVTLSLKVGRCVYDDAHAAAGRDGFVAGRRAEFDSADG 118
QY	102 VRKSIDESLERLQDLYVDILHCHDI-----EFGSLDQIVSETIPALQKLQEGKTRF 153
Dd	119 VRRAPASSIGRGTDYIDVLILLHDIGALTGDHNANLVQLAEELPA MAELKAACGCA 178
QY	154 IGTGLPLDIFTYVLDVRPPGTVDVLTSCHYGVNDSTLLD-----LIPYLKSKGVGVI 207
Dd	179 IGLGVNEQDVALEVLPRFF-----LDCVMLAGRY-----TLLEQHGA RALLQQAAQRNVAIL 230
QY	208 SASPLAWGLLTE-QGPPEWHPASP----ELKSASKAAVAHCSSKGKKITKLALOYSIA NK 262
Dd	231 SAGPYSSGGLSDARGPGATYNYPVDVTATLQAQRLYAA-CAAFGV DIGAALLOFLAHP 289
QY	263 EISSVLGVNSSVQVEENVAAVTELESGLMDQBTLSERVEBAILEPVKNLTW 312
Dd	290 AVTTVVAGNRSAEVNR---SAASTRLLA-----PVP SALW 320
RESULT 12	
Q9VGF2	
ID	PRELIMINARY; PRT; 294 AA.
AC	Q9VGF2
CD	01-MAY-2000 (TReMBLrel. 13, Created)
DT	01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT	01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE	CG12224-PA.
GN	ORFNames=CG12224;
OS	Drosophila melanogaster (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;	
[1]	
SEQUENCE FROM N.A.	
MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;	
Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,	
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
Sutton G.G., Wortman J.R., Yeaman M.D., Zhang Q., Chen L.X.H.C.,	
Brandon R.C., Rogers J.H., Blazej R.G., Champe M., Pfeiffer B.D.,	
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabo G.L.,	
Abri J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,	
Baillet R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,	
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,	
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,	
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,	
Foster C., Gabrielian A.S., Farg N.S., Gelbart W.M., Glasser K.,	
Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,	
Harris N.L., Harvey D., Heintz T.J., Hernandez J.R., Houck J.,	
Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,	
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,	
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,	
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,	
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,	
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefia,	

```

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.W., Weissbach J.,
RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genome perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
[4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Mirza S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
[5]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
[6]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003694; AAF54730.1; -
DR FlyBase; FBGN0037974; CG12224.
DR InterPro; IPR001395; Aldol/ket_red.
DR Pfam; PF00248; Aldol_ket_red; I.
DR PRINTS; PR00069; ALDKETDTHASE.
SQ SEQUENCE 294 AA; 32984 MW; 8B03E0F0C8EE1B8E CRC64;

Query Match 26.0%; Score 420.5; DB 2; Length 294;
Best Local Similarity 33.7%; Pred. No. 5.9e-22;
Matches 102; Conservative 57; Mismatches 101; Indels 43; Gaps 7;

QY 3 KIELRALGNTGLKSAVGFSGPLSGVFGVAEDDAVATVREAPRLGNFDFTSPPYGGT 62
DB :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 21 RMEYRQLAGTGLHVKSLAIGGSPLCNL-----FFDDYDREGI 58
QY 63 LSEKMLGKGLKALQVPRSDYIVATKCGRY-----KEGDFPSAERVKSIDESLERLQDYY 118

```

```

Db 59 L-----MALKDVPREAYIATKVARYGLDPKQMFEDYSADKARESVKRSLERLQDRV 110
QY 119 DILCHDIEFG-SLDQIVSETIPALQKLKQEGKTRFTGITGLPDLITVYLDVRPPGTVD 177
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 111 DILQVHDVAAPNDLIVLNETIPVLEEVQAGKARFIVGTAYDVLKCEAER-GKGRIQ 169
QY 178 VILSYCHYGVNDSTLLDLLLPYLSKGVGVISASPLANGLLTEQGPPEWHPASPELKSASK 237
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 170 VLVNARYTLLDNTLLRYMKDFQMGVGVVCAAAHSLGLLRNAGHSHPGSQBILAVAK 229
QY 238 AAVACHSKGKIKTKALQYSLANKETSSVLVGVSSVSQVENVAAVTELESGLM---DQ 294
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 230 RGABICQORNVLEKGLAMTYTMDLGAATFLIGIPNKLRLINLDAIFD----GLTSHEQ 285
QY 295 ETL 297
Db :||
Db 286 EVL 288

RESULT 13
QSP3K6
ID QSP3K6 PRELIMINARY; PRT; 335 AA.
AC Q8P3K6;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DN Oxidoreductase.
GN OrderedLocusNames=XCC4065;
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217; DOI=10.1038/417459a;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
RA Cimarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.W., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL; AE012528; AAM43286.1; -
DR InterPro; IPR001395; Aldol/ket_red.
DR Pfam; PF00248; Aldol_ket_red; 1.
DR ProDom; PD000288; Aldol/ket_red; 1.
DR Complete proteome.
SQ SEQUENCE 335 AA; 34773 MW; 36B212C16AFA2F1F CRC64;

Query Match 25.8%; Score 417.5; DB 2; Length 335;
Best Local Similarity 35.2%; Pred. No. 1.1e-21;
Matches 108; Conservative 54; Mismatches 98; Indels 47; Gaps 9;

QY 9 LGNTGLKSAVGFSGPLSGVFGVAEDDAVATVREAPRLGNFDFTSPPYGGTISEKVL 68
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 8 VGSTAVQSLSAGFGAIPGNIYSVDEGDAALAVTGAFAEIRHFDAPFYGYGLSEARL 67
QY 69 GKGLKALQVPRSDYIVATKCGRY-----KEG-----FDFSAERVKSIDES 109
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 68 GRGLAG--VPRADYTLTKVGRVCVYDDATAVAGRDGFAVAGRAEFDYSAANGVRAPASS 125

```


Db 145 VTTLHEMRKEGKIHYVGISGYPVDFLAWCEILRDQVGP---VDIVLSYSNNCLQNTLLL 201
QY 194 DLLPYLKSGV-GVISASPLAMGLLTEQGPPEWHFASPELKSASKAHVAHCKSKGKITK 252
Db 202 ESLDRFKLAGVQKVLNGSPLSLRSQPTHSFHAPESLNKAVSKVVEYTKSQDVELAD 261
QY 253 LALQYSLANKEISSVLVGMSSVSQVEENVA-----VTEBSLGMDOETLSEVEAILEPV 307
Db 262 LALRYAFA-KPPGTVVG---VSTVEELQAALDNYWLAKSESVS-DOEMFONVKSLYGDQ 316
QY 308 KNLTPWPSG 315
Db 317 WNLWTWTS 324

Search completed: November 13, 2005, 08:38:50
Job time : 85 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 13, 2005, 07:01:27 ; Search time 112.5 Seconds
(without alignments)
8739.485 Million cell updates/sec

Title: US-10-606-300-12

Perfect score: 1653

Sequence: 1 atgcgaaaatagagcttcg.....gtggatccatcagaactaa 960

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool_h/US10606300/runat_07112005_094818_7301/app_query.fasta_1.1159
-DB=UniProt -QFMT=fastcan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOPCPU=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10606300@cgn 1 1 106 @runat_07112005_094818_7301 -NCPU=6 -ICPU=3
-NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt 03.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1619	97.9	319	081884	081884 arabidopsis
2	1348	81.5	319	Q84L11	Q84L11 actinidia d
3	1322	80.0	324	Q84L20	Q84L20 malus domes
4	1261	76.3	322	Q6BDJ2	Q6BDJ2 spinacia ol
5	695	42.0	310	Q8A1P5	Q8A1P5 bacteroides
6	625	37.8	347	Q7U2A0	Q7U2A0 rhodospirall
7	593.5	35.9	345	Q9VGF3	Q9VGF3 drosophila
8	593	35.9	342	Q7QHC3	Q7QHC3 anopheles g
9	544.5	32.9	342	Q9VGF1	Q9VGF1 drosophila
10	503	30.4	439	Q20127	Q20127 caenorhabdi
11	426	25.8	336	Q8PF04	Q8PF04 xanthomonas
12	420.5	25.4	294	Q9VGF2	Q9VGF2 drosophila
13	417.5	25.3	335	Q8PFK6	Q8PFK6 xanthomonas
14	414.5	25.1	336	Q6FU59	Q6FU59 candida gla
15	413	25.0	330	Q6BZU5	Q6BZU5 yarrowia li
16	387.5	23.4	359	Q75DD3	Q75DD3 ashbya goss

17	383	23.2	328	2	Q82N99	Q82N99 streptomyce
18	380	23.0	433	2	Q7S6W9	Q7S6W9 neurospora
19	375	22.7	328	2	Q6CIG5	Q6CIG5 kluyveromyc
20	369.5	22.4	328	2	Q826A9	Q826A9 streptomyce
21	366	22.1	329	2	Q52472	Q52472 pseudomonas
22	362.5	21.9	297	2	Q9KE47	Q9KE47 bacillus ha
23	361	21.8	325	2	Q89VC3	Q89VC3 bradyrhizob
24	358	21.7	335	1	YMT1_YEAST	Q04212 saccharomyc
25	349	21.1	306	2	Q65HS3	Q65HS3 bacillus li
26	346.5	21.0	329	2	Q8UA47	Q8UA47 agrobacteri
27	346.5	21.0	434	2	Q7XEJ9	Q7XEJ9 oryza sativ
28	345	20.9	369	2	Q6BPT1	Q6BPT1 debaryomyce
29	344.5	20.8	334	2	Q9RJW1	Q9RJW1 streptomyce
30	342.5	20.7	326	2	Q7NMCH8	Q7NMCH8 gloeobacter
31	342.5	20.7	329	2	Q8ETF4	Q8ETF4 oceanobacil
32	339	20.5	304	2	Q81820	Q81820 bacillus ce
33	337.5	20.4	306	1	YQKP_BACSU	P54569 bacillus su
34	336.5	20.4	314	2	Q7NK63	Q7NK63 gloeobacter
35	329.5	19.9	304	2	Q81MD1	Q81MD1 bacillus an
36	329.5	19.9	304	2	Q6HE69	Q6HE69 bacillus th
37	327.5	19.8	352	2	Q6F726	Q6F726 acinetobact
38	327	19.8	352	2	Q6DAK3	Q6DAK3 erwinia car
39	325.5	19.7	304	2	Q731K1	Q731K1 bacillus ce
40	323.5	19.6	304	2	Q635I5	Q635I5 bacillus ce
41	323.5	19.6	339	2	Q92TD8	Q92TD8 rhizobium m
42	317	19.2	333	2	Q64UL3	Q64UL3 bacteroides
43	317	19.2	351	2	Q8UAD6	Q8UAD6 agrobacteri
44	315	19.1	326	2	Q8FGX5	Q8FGX5 escherichia
45	314	19.0	325	2	Q8G5A3	Q8G5A3 bifidobacte

ALIGNMENTS

RESULT 1

081884 PRELIMINARY; PRT; 319 AA.
ID 081884;
AC 081884;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Hypothetical protein T16L1.160 (AT4933670/T16L1.160) (L-galactose
DE dehydrogenase) (Hypothetical protein AT4933670).
GN Name=T16L1.160; Synonyms=AT4933670, L-GalDH;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids;
OC eucotids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Obermaier B., Deutschenbaur S., Piravandi E., Hoheisel J.,
RA Jesse T., Heijnen L., Vos P., Meves H.W., Mayer K.F.X., Schueller C.,
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Koesena E., Chen H., Cheuk R., Kim C.J., Meyers M.C., Shinn P.,
RA Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,

RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=22043773; PubMed=12047629;
RA Gatzek S., Wheeler G.L., Smirnov N.;
RT "Antisense suppression of l-galactose dehydrogenase in Arabidopsis
RT thaliana provides evidence for its role in ascorbate synthesis and
RT reveals light modulated l-galactose synthesis.";
RL Plant J. 30:541-553(2002).
RN [6]
RP SEQUENCE FROM N.A.
RA Obermaier B., Deutschenbaur S., Piravandi E., Mewes H.W., Lemcke K.,
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031394; CAA20580.1; -;
DR EMBL; AY050377; AAK91395.1; -;
DR EMBL; AY090337; AAL90998.1; -;
DR EMBL; AJ417563; CAD10386.1; -;
DR EMBL; AL161583; CAB80084.1; -;
DR PIR; T04984; T04984.
DR HSSP; P52895; IJ396.
DR InterPro; IPR001395; Aldo/ket_red.
DR Pfam; PF00248; Aldo_ket_red; 1.
DR PRINTS; PR00069; ALDKETRDASE.
DR ProDom; PD000288; Aldo/ket_red; 1.
KW Hypothetical protein.
SQ SEQUENCE 319 AA; 34531 MW; 9400717380DFA71E CRC64;

Alignment Scores:
Pred. No.: 1.4e-117 Length: 319
Score: 1619.00 Matches: 319
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 97.94% Indels: 0
DB: 2 Gaps: 0

US-10-606-300-12 (1-960) x 081884 (1-319)

QY 1 ATGCGAATAATAGCTTCGAGCTTTGGGACACAGGCTTAAGGTAGCGCGTTCGT 60
DB 1 MetThrLysIleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGly 20

QY 61 TTTGGTGCCTCCGCTCGGAAGTGTCTTCGGTCCAGTCCGCGAAGATGATCGCCGCGC 120
DB 21 PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAla 40

QY 121 ACCGTGCGCGAGCTTCGCTCCGGTATCAACTTTCGACACCTCCCGATTATATGGA 180
DB 41 ThrValArgGluAlaPheArgLeuGlyIleAsnPheAspThrSerProTyrTyrGly 60

QY 181 GGAACTCTCTGAGAAATGCTTGGTAGGGAGCTTAAGGCTTCGAGTCCCTAGAGT 240
DB 61 GlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSer 80

QY 241 GACTACATTGGCTACTAAGTGTGGTAGATATAAAGAGGTTTTGATTTCAAGTCTGAG 300
DB 81 AspTyrIleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGlu 100

QY 301 AGAGTAAGAAGAGTATTACCGAGAGCTTTGGAGAGCTTCAGCTGATTTATTCACATA 360
DB 101 ArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 120

QY 361 CTTTCATGCGCATGACATTGAGTTCGGGTCTCTTGATCAGATTGTGATGAACAAATTCCT 420

DB 121 LeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePro 140
QY 421 GCTCTTCAGAACTGAACAGAGGGAAGACCGGTTTCATTGTGTATCAGTGTCTCCG 480
DB 141 AlaLeuGlnLysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGlyLeuPro 160
QY 481 TTAGATATTTTCACTTATGTTCTTGATCGAGTGCCTCAGGAGCTGTCTGATGTGATTTG 540
DB 161 LeuAspIlePheThrThrValLeuAspArgValProGlyThrValAspValIleLeu 180
QY 541 TCATACCTGCTATTACGGGTTAATGATTCGACGTTGCTGGATTTACTACCTTACTGAAG 600
DB 181 SerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLys 200
QY 601 AGCAAGAGTGTGGGTGTGATAAGTGTCTTCATAGCAATGAGGCTCTCTTACAGAACAA 660
DB 201 SerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuThrGluGln 220
QY 661 GGTCTCTGTAATGGCACCTGCTTCCCTCAGCTCAAGTCTGCAAGCAAGCCGCGCTT 720
DB 221 GlyProGluTrpHisProAlaSerProGluLeuLysSerAlaSerLysAlaVal 240
QY 721 GCTCACTGCAATCAAAAGGCAAGATCAAAAGTTAGCTCTGCAATACAGTTTAGCA 780
DB 241 AlaHisCysLysSerLysGlyLysIleThrLysLeuAlaLeuGlnTyrSerLeuAla 260
QY 781 AACAGAGAGATTTCGTGCTGCTGGTGGATGAGCTCTGCTCAGAGTGAAGAAAT 840
DB 261 AsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGlnAsn 280
QY 841 GTTCGACAGTTACAGAGCTTGAAGTCTGGGATGATCAAGAACTCTCTCTGAGTT 900
DB 281 ValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluVal 300
QY 901 GAAGCTATTCTCGAGCCTGTAAAGAACTGATGATGCGCAAGTGAAGTGAATCCATCAGAAC 957
DB 301 GluAlaIleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGlnAsn 319

RESULT 2
Q84LI1 PRELIMINARY; PRT; 319 AA.
AC Q84LI1;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Galactose dehydrogenase.
OS Actinidia deliciosa (Kiwifruit).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC Ericales; Actinidiaceae; Actinidia.
RX NCBI_TaxID=3627;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Developing shoot buds;
RA Beuning L., Bowen J., Crowhurst R., Gleave A., MacRae E., Newcomb R.,
RA Perera S., Ross G., Snowden K., Walton E., Yauk Y.K.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Developing shoot buds;
RA Laing W.A., MacRae E.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY176585; AAO18639.1; -;
DR InterPro; IPR001395; Aldo/ket_red.
DR Pfam; PF00248; Aldo_ket_red; 1.
DR PRINTS; PR00069; ALDKETRDASE.
DR ProDom; PD000288; Aldo/ket_red; 1.
SQ SEQUENCE 319 AA; 34531 MW; 8657434C03520FAB CRC64;

Alignment Scores:
Pred. No.: 1.88e-96 Length: 319
Score: 1348.00 Matches: 256
Percent Similarity: 90.57% Conservative: 32

Best Local Similarity: 80.50% Mismatches: 30
Query Match: 81.55% Indels: 0
DB: 2 Gaps: 0

US-10-606-300-12 (1-960) x Q84L20 (1-319)

Qy 1 ATGACGAAATAGAGCTTCGAGCTTGGGGACACAGGGCTTAAGTTAGCGCGTGGT 60
Db 1 MetThrLeuAspLeuArgProLeuGlyAsnThrGlyLeuLysLeuSerValGly 20
61 TTTGGTGTCTCGCTCGGAGTCTCTCGTCCAGTCGCGAGAGATGATCGCTGCC 120
Db 21 PheGlyAlaSerProLeuGlyAsnValPheGlyProValSerAspHisAlaLeu 40
121 ACCGTGCGCGAGCTTCCGCTCGGTATCACTTCTTCGACACCTCCCGTATTATGGA 180
Db 41 SerValArgGluAlaLeuArgLeuGlyIleAsnPheAspThrSerProTyrGly 60
191 GGAACACTGTCTGAGAAATGCTTGGTAAGGACTAAAGGCTTTCGAAGTCCCTAGAGT 240
Db 61 GlyThrLeuSerGluLysValLeuGlyLysAlaLeuLysAlaLeuGlyValProArgAsn 80
241 GACTACATTTGGCTTACTAAGTGTGTAGATATATAAGAGTCTTTGATTTTCAGTCTGAG 300
Db 81 GluTyrIleValSerThrLysCysGlyArgTyrAlaGluGlyPheAspPheSerAlaGlu 100
301 AGAGTAAGAAAGATGATTGACGAGAGCTTGAGAGGCTTCAGCTTGATTTATGTTGACATA 360
Db 101 ArgValThrLysSerPheAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 120
361 CTTCACTGCCATGACATTGAGTTCGGTCTCTTGATCAGATGTGAGTGAACAATTCCT 420
Db 121 LeuGlnCysHisAspIleGluPheGlySerLeuAspGlnIleValAsnGluThrIlePro 140
421 GCTCTTCAGAACTCAACAGAGGGAGACCGGTTCACTTGTGATCCTGCTGCTCCG 480
Db 141 AlaLeuGlnLysLeuLysGluAlaGlyLysIleArgPheIleGlyIleThrGlyLeuPro 160
481 TTAGATATTTTCACTTATGTTCTTCATCGAGTGCCTCCAGGAGCTGTCGATGTGATTTG 540
Db 161 LeuGlyValPheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeu 180
541 TCATCTCTCATTTACGGCTTAAGTTCGACGTTGCTGGATTTACTACTTACTTTGAAG 600
Db 181 SerTyrCysHisTyrSerIleAsnAspSerThrLeuGluAspLeuLeuProTyrLeuLys 200
601 AGCAAGTGTGGTGTGATAGTCTTCTCCATAGCAATGGGCTCTCTACAGAACAA 660
Db 201 SerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluSer 220
661 GTCTCTCGTAATGCGCACCTGCTTCCCTCGAGCTCAAGTCTGCAAGCAAGCCGAGTT 720
Db 221 GlyProGluThrPheProAlaSerProGluLeuLysAlaAlaCysGlnAlaAla 240
721 GCTCATCTCAATCAAGCGCAAGAGATCAAAAGTTAGCTTCTGCAATACAGTTTAGCA 780
Db 241 AlaHisCysLysGluLysGlyArgAsnIleSerLysLeuAlaMetGlnTyrSerLeuSer 260
781 AACAGGAGATTTCTCGGTGTGGTGTGGATGAGCTCTGCTCTACAGTAGAGAAAT 840
Db 261 AsnLysAspIleSerSerIleLeuValGlyMetAsnSerValLysGlnValGluGluAsn 280
841 GTTGACAGATTTACAGAGCTGAAGTCTGGGATGATCAAGAACTCTGTGTGAGTT 900
Db 281 ValAlaAlaAlaAsnGluLeuAlaThrPheGlyLysAspGluLysThrValSerGluIle 300
901 GAAGCTATTCTCGACCTGTAAGAAATCTGACATGCCCAAGTGAATCCATCAG 954
Db 301 GluGluIleLeuLysProValLysAsnGlnThrTrpLeuSerGlyIleGlnGln 318

Qy 1 ATGACGAAATAGAGCTTCGAGCTTGGGGACACAGGGCTTAAGTTAGCGCGTGGT 60
Db 6 LeuProLysValGluLeuArgGluLeuGlyAsnThrGlyLeuLysLeuSerCysValGly 25
61 TTTGGTGTCTCGCTCGGAGTGTCTTCGGTCCAGTCGCGAGAGATGATCGCTGCC 120
Db 26 PheGlyAlaSerProLeuGlyAsnValPheGlyProValSerAspGluAlaIleGly 45
121 ACCGTGCGCGAGCTTCCGCTCGGTATCACTTCTTCGACACCTCCCGTATTATGGA 180
Db 46 SerValArgGluAlaPheArgGlyIleAsnPheAspThrSerProTyrTyrGly 65
181 GGAACACTGTCTGAGAAATGCTTGGTAAGGAGCTTAAAGGCTTTCGAAGTCCCTAGAAAT 240
Db 66 GlyThrLeuSerGluLysValLeuGlyLysThrLeuLysAlaLeuGlyValProArgSer 85
241 GACTACATTTGGCTTACTAAGTGTGTAGATATATAAGAGGCTTTCAGTCTGAG 300
Db 86 GluTyrValValAlaThrLysCysGlyArgTyrAlaAspGlyPheAspPheSerAlaAsp 105
301 AGAGTAAGAAAGATGATTGACGAGAGCTTGAGAGGCTTTCAGTGTGATTTATGTTGACATA 360
Db 106 ArgValThrLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 125
361 CTTCACTGCCATGACATTGAGTTCGGTCTCTTGATCAGATGTGAGTGAACAATTCCT 420
Db 126 LeuGlnCysHisAspIleGluPheGlySerLeuAspGlnIleValAsnGluThrIlePro 145
421 GCTCTTCAGAACTCAACAGAGGGAGACCGGTTCACTTGTGATTCAGTCTTCCG 480
Db 146 AlaLeuLysLysLeuLysGluAlaGlyLysIleArgPheIleGlyIleThrGlyLeuPro 165
481 TTAGATATTTTCACTTATGTTCTTCATCGAGTGCCTCCAGGAGCTGTGATGTGATTTG 540

Alignment Scores:
Pred. No.: 2,01e-94 Length: 324
Score: 1322.00 Matches: 248
Percent Similarity: 90.28% Conservative: 40
Best Local Similarity: 77.74% Mismatches: 31
Query Match: 79.98% Indels: 0
DB: 2 Gaps: 0

US-10-606-300-12 (1-960) x Q84L20 (1-324)

Qy 1 ATGACGAAATAGAGCTTCGAGCTTGGGGACACAGGGCTTAAGTTAGCGCGTGGT 60
Db 6 LeuProLysValGluLeuArgGluLeuGlyAsnThrGlyLeuLysLeuSerCysValGly 25
61 TTTGGTGTCTCGCTCGGAGTGTCTTCGGTCCAGTCGCGAGAGATGATCGCTGCC 120
Db 26 PheGlyAlaSerProLeuGlyAsnValPheGlyProValSerAspGluAlaIleGly 45
121 ACCGTGCGCGAGCTTCCGCTCGGTATCACTTCTTCGACACCTCCCGTATTATGGA 180
Db 46 SerValArgGluAlaPheArgGlyIleAsnPheAspThrSerProTyrTyrGly 65
181 GGAACACTGTCTGAGAAATGCTTGGTAAGGAGCTTAAAGGCTTTCGAAGTCCCTAGAAAT 240
Db 66 GlyThrLeuSerGluLysValLeuGlyLysThrLeuLysAlaLeuGlyValProArgSer 85
241 GACTACATTTGGCTTACTAAGTGTGTAGATATATAAGAGGCTTTCAGTCTGAG 300
Db 86 GluTyrValValAlaThrLysCysGlyArgTyrAlaAspGlyPheAspPheSerAlaAsp 105
301 AGAGTAAGAAAGATGATTGACGAGAGCTTGAGAGGCTTTCAGTGTGATTTATGTTGACATA 360
Db 106 ArgValThrLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 125
361 CTTCACTGCCATGACATTGAGTTCGGTCTCTTGATCAGATGTGAGTGAACAATTCCT 420
Db 126 LeuGlnCysHisAspIleGluPheGlySerLeuAspGlnIleValAsnGluThrIlePro 145
421 GCTCTTCAGAACTCAACAGAGGGAGACCGGTTCACTTGTGATTCAGTCTTCCG 480
Db 146 AlaLeuLysLysLeuLysGluAlaGlyLysIleArgPheIleGlyIleThrGlyLeuPro 165
481 TTAGATATTTTCACTTATGTTCTTCATCGAGTGCCTCCAGGAGCTGTGATGTGATTTG 540

PRELIMINARY; PRT; 324 AA.

RESULT 3
Q84L20
ID Q84L20

Db 166 LeuGlyIlePheThrTyrrValLeuAspArgIleProAlaGlyMetValaspValValLeu 185
QY 541 TCATACCTGTCATACGGGCTTAATGATTCGAGCTTGGATTTACTACCTTACTTGAAG 600
Db 186 SerTyrrCysHisTyrrGlyIleAsnAspSerThrLeuGluAspLeuProTyrrLeuLys 205
QY 601 AGCAAGGTGGGTGTGATAGTCTCTCCATAGCAATGGGCTCTCTACAGAACAA 660
Db 206 SerLysGlyValGlyIleSerAlaSerProLeuAlaMetGlyLeuThrGluAsn 225
QY 661 GTCCTCTGTAAGTGCACCTCTCTCCCTGAGCTCAAGTCTGCAAGCAAGCGCAGTT 720
Db 226 GlyProGluTrpHisProAlaSerAlaGluLeuLysSerAlaCysArgAlaAla 245
QY 721 GCTCACTGCAATCAAGGGCAAGAGATCAAAAGTTAGTCTCTCAATACAGTTTAGCA 780
Db 246 ValTyrrCysLysGluArgGlyAsnAsnIleSerLysLeuAlaLeuGlnTyrrSerLeuSer 265
QY 781 AACAGGAGATTTCTCGGTGTGGTGGATGAGCTCTGCTCACAGGTAGAGAAAT 840
Db 266 AsnLysAspIleSerSerValLeuValGlyMetAsnSerIleAsnGlnValGluGluAsn 285
QY 841 GTTCAGCAGTTACAGAGCTGAAGTCTGGGATGGATCAAGAACTCTGCTCAGGTT 900
Db 286 ValAlaAlaValGluLeuAlaThrIleGlyLysAsnGluLysIleLeuAlaGluVal 305
QY 901 GAAGCTATTCTCGACCTGTAAGAACTGACATGCGCAAGTGGATCCATCAGAAC 957
Db 306 GluAlaIleLeuLysProValLysAsnGlnThrTrpProSerGlyLeuGlnGlnSer 324

RESULT 4

Q6BDJ2 PRELIMINARY; PRT; 322 AA.
AC Q6BDJ2;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE L-galactose dehydrogenase.
GN Name=GDH;
OS Spinacia oleracea (Spinach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
OC Caryophyllales; Amaranthaceae; Spinacia.
OX NCBI_TaxID=3562;
RN [1]
RP SEQUENCE FROM N.A.
RA Mieda T., Yabuta Y., Rapolu M., Motoki T., Ishikawa T., Yoshimura K.,
RA Shigeoka S.;
RT "Direct Submission."
RL Submitted (JAN-2004) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB160990; BAD32687.1; -.
DR InterPro; IPR001395; Aldo/ket red.
DR Pfam; PF00248; Aldo_ket_red; 1.
DR PRINTS; PR00089; ALDKETRDASE.
DR ProDom; PD000288; Aldo/ket red; 1.
SQ SEQUENCE 322 AA; 35262 MW; CB5BA18D81A27AF4 CRC64;

Alignment Scores:

Pred. No.: 1.14e-89 Length: 322
Score: 1261.00 Matches: 237
Percent Similarity: 87.03% Conservative: 38
Best Local Similarity: 75.00% Mismatches: 41
Query Match: 76.29% Indels: 0
DB: 2 Gaps: 0

US-10-606-300-12 (1-960) x Q6BDJ2 (1-322)

QY 7 AAATAGAGCTTCGAGCTTTGGGGAACACAGGCTTAAGGTTAGCGCGCTGGTTTGGT 66
Db 6 LysLeuGluArgGluLeuGlyAsnThrGlyLeuAsnLeuSerCysValGlyPheGly 25
QY 67 GCCTCTCGCTCGGAAGTGTCTTCGGTCCAGTCGCGGAGATGATGCGCCACCGTG 126
|||||

Db 26 AlaSerProLeuGlyAsnValPheGlyAspValSerGluGluGlnSerIleAlaThrVal 45
QY 127 CGCAGGCTTTCCGCTCTCGGTATCAACTTTCTCGACACCTCCCGGTATTATGGAGAA 186
Db 46 IleGluAlaPheAsnGlnGlyIleAsnPhePheAspThrSerProTyrrTyrrGlyAlaThr 65
QY 187 CTGTCTGAGAAATGCTTGGTAAAGGACTAAAGCTTTGCAAGTCCCTAGAGTGACTAC 246
Db 66 LeuSerGluLysValLeuGlyLysCysLeuLysAlaLeuGlyAlaSerArgAspGluTyrr 85
QY 247 ATTGTGGCTACTAGTGTGGTAGATATAAAGAGGTTTTTGATTTTCAGTGTGAGAGAT 306
Db 86 IleValAlaThrLysCysGlyArgTyrrAlaGluGlyPheAspPheSerAlaGluArgVal 105
QY 307 AGAAAGAGTATTGACGAGAGCTTGGAGAGCTTTCAGCTTGATTTGTTGACATCTTCA 366
Db 106 ThrLysSerIleAspGluSerLeuGluArgLeuGluGlnGluTyrrValAspIleLeuGln 125
QY 367 TGCCATGACATTCAGTTCGGGTCTCTTCATCAGATTGTGAGTGAACAATTCCTGCTCT 426
Db 126 CysHisAspIleGluPheGlySerLeuAspGlnIleValAsnGluThrIleProAlaLeu 145
QY 427 CAGAACTGAAACAGAGGGGAGACCGGTTTCATTGTGTATCCTCTGCTCTCCGTTAG 486
Db 146 GlnLysIleLysGluSerGlyLysThrArgPheIleGlyIleThrGlyLeuProLeuGln 165
QY 487 ATTTTCATTATGTTCTTTCGATCGAGTGCCTCCAGGAGCTTCGATGTGATATTGTCA 546
Db 166 ValTyrrThrTyrrValLeuAspArgValProProGlyThrIleAspValValLeuSerTyrr 185
QY 547 TGTCAATACGGCGTTAATGATTCGACGCTTCGATTCGATTTACTACCTTACTTGAAGACA 606
Db 186 CysHisTyrrCysIleAsnAspSerThrLeuGluAspMetLeuProTyrrPheLysSerLys 205
QY 607 GGTCTGGGTGATAGTTCCTCCATTGCAATGGGCTTCCTTACAGAACAAAGTCTCT 666
Db 206 GlyValGlyValIleAsnAlaSerProLeuSerMetGlyLeuHisThrGluAsnGlyPro 225
QY 667 CCTGAATGGCACCTCTCTCCCTCGAGTCTCAAGTCTGCAAGCAAGCCGCGAGTTCTCAC 726
Db 226 ProGluTrpHisProAlaSerProGluLysAlaAlaCysLysAlaAlaAspTyrr 245
QY 727 TGCATAATCAAGGGCAGAGATCAAAAGTTAGTCTGCAATCAGTTTACAGTTTACCAAC 786
Db 246 CysLysLysAsnGlyLysAsnIleSerLysLeuAlaLeuGlnTyrrSerLeuSerAsnLys 265
QY 787 GAGATTTCTCGTCTGGTGTGGTGGATGAGTCTGTCTCACAGGTAGAAGAAATCTTGCA 846
Db 266 AspIleSerThrThrLeuValGlyMetAsnSerValLysGlnValGluGluAsnValGly 285
QY 847 GCAGTTACAGAGCTTGAAGTCTGGGATGGATCAAGAACTCTGTCTGAGGTTGAAGCT 906
Db 286 AlaAlaLeuGluLeuGluThrAlaGlyLysAspGluLysThrPheAlaGluIleGluAsn 305
QY 907 ATTCTCGAGCTCTAAAGAACTGACATGGCCCAAGTGAATCCATCAG 954
Db 306 IleLeuLysProIleLysAsnGlnSerTrpProSerGlyIleGlnGln 321

RESULT 5

Q8A1P5 PRELIMINARY; PRT; 310 AA.
ID Q8A1P5;
AC Q8A1P5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative oxidoreductase.
GN OrderedLocustNames=BT3614;
OS Bacteroides thetaiotaomicron.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=818;
RN [1]
RP SEQUENCE FROM N.A.

```
RC STRAIN=VPI-5482 / ATCC 29148;
RX MEDLINE=22550858; PubMed=12663928; DOI=10.1126/science.1080029;
RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RA Chiang H.C., Hooper L.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
RL Science 239:2074-2076(2003).
DR EMBL; AEO16941; AAO78719.1; -.
DR InterPro; IPR001395; Aldo_ket_red.
DR Pfam; PF00248; Aldo_ket_red; 1.
DR PRINTS; PR00069; ALDKETRDASE.
DR ProDom; PD000288; Aldo/ket_red; 1.
KW Complete proteome.
SQ SEQUENCE 310 AA; 34831 MW; 2D6C1B689C1C5298 CRC64;

Alignment Scores:
Pred. No.: 1.52e-45 Length: 310
Score: 695.00 Matches: 135
Percent Similarity: 66.45% Conservative: 73
Best Local Similarity: 43.13% Mismatches: 95
Query Match: 42.04% Indels: 10
DB: 2 Gaps: 4

US-10-606-300-12 (1-960) x Q8A1P5 (1-310)
QY 10 ATAGAGCTTCGAGCTTTGGGGAACACAGGCTTAAGTTAGCGCGTGGTGGTGGTCC 69
DB 1 MetGlnTyrHisGluIleGlyThrGlyMetLysValSerSerLeuSerPheGlyAla 20
QY 70 TCTCCGCTCGGAAGTCTTCGGTCCAGTCCGCGAAGATGATCGCGCCACCGTGGC 129
DB 21 SerSerLeuGlyGlyValPheHisAspLeuLysGluGlyIleGlnAlaValPhe 40
QY 130 GAGGCTTTCCGCTCTCGGTATCAACTCTTCGACACCTCCCGGTATATGGAGGAACACTG 189
DB 41 ThrAlaIleGluAlaGlyMetAsnPheIleAspValSerProTyrTyrGlyHisTyrLys 60
QY 190 TCTGAGAAATGCTTGGTAAGGACTAAAGCTTTGCAAGTCCCTAGAGTGACTACATT 249
DB 61 AlaGluThrValLeuGlyLysAlaLeuLys-----AspIleProArgAspArgTyrTyr 78
QY 250 GTGGCTACTAGTGTGGTAGATAT---AAGNAGGT-----TTTGATTTCAGTGCT 297
DB 79 LeuSerThrLysValGlyArgTyrGlyLysAspGlyValAsnThrTyrAspTyrSerAla 98
QY 298 GAGAGGTAAGAAAGATTTGACGAGAGCTTGAGAGGCTTCAGCTTGATTATGTTGAC 357
DB 99 LysArgAlaThrGluSerValTyrGluSerMetGluArgLeuAsnIleAspPheIleAsp 118
QY 358 ATACTTCATGCCATGACATTCAGTTCGGGTCTCTTGATCATGATTCAGTGGAGAACAAATT 417
DB 119 LeuIleAsnValHisAspIleGluPheAlaAspLeuAsnGlnValValAsnGluThrLeu 138
QY 418 CTTGCTCTTCAGAACTGAAACAGAGGGGAGACCGGTTTCATTGGTATCATCTGCTT 477
DB 139 ProAlaLeuValGluLeuArgGluLysGlyValValGlyHisValGlyIleThrAspLeu 158
QY 478 CCGTTAGATATTTTCACTATGTTCTTCATCGAGTCCCTCCAGGACCTGTCGATGTGATA 537
DB 159 GlnLeuGluAsnLeuLysTyrPheValIleAspArgSerProSerGlyThrIleGluSerVal 178
QY 538 TTGTCTACTGTCAATACGGGCTTAATGATTCGAGCTTGGTGGATTTACTACCTTACTTG 597
DB 179 LeuSerPheCysHisTyrCysLeuLysAspAspLysLeuAlaAspPheLeuAspTyrPhe 198
QY 598 AAGACCAAGGTGGGTGTGATAGTCTTCTTCATTAGCATGGGCTCTTACAGAA 657
DB 199 GluSerLysGluIleGlyValIleAsnAlaSerProLeuSerMetGlyLeuSerGlu 218
QY 658 CAAGTCTCTCCTGAATGGCACCTCTCTCCCTGAGCTCAAGTCTGCAAGCAAGCCGCA 717
DB 219 ArgGlyValProValTyrPheIleAsnAlaProLysProLeuValAspAlaCysArgLysAla 238
QY 718 GTTGCTCACTGCAATCAAGGGCAAGAGATCACAAAGTTAGTCTCTGCAATACAGTTTA 777

STRAIN=VPI-5482 / ATCC 29148;
MEDLINE=22550858; PubMed=12663928; DOI=10.1126/science.1080029;
Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
Chiang H.C., Hooper L.V., Gordon J.I.;
"A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
Science 239:2074-2076(2003).
EMBL; AEO16941; AAO78719.1; -.
InterPro; IPR001395; Aldo_ket_red.
Pfam; PF00248; Aldo_ket_red; 1.
PRINTS; PR00069; ALDKETRDASE.
ProDom; PD000288; Aldo/ket_red; 1.
Complete proteome.
SEQUENCE 310 AA; 34831 MW; 2D6C1B689C1C5298 CRC64;

Alignment Scores:
Pred. No.: 1.52e-45 Length: 310
Score: 695.00 Matches: 135
Percent Similarity: 66.45% Conservative: 73
Best Local Similarity: 43.13% Mismatches: 95
Query Match: 42.04% Indels: 10
DB: 2 Gaps: 4

US-10-606-300-12 (1-960) x Q8A1P5 (1-310)
QY 10 ATAGAGCTTCGAGCTTTGGGGAACACAGGCTTAAGTTAGCGCGTGGTGGTGGTCC 69
DB 1 MetGlnTyrHisGluIleGlyThrGlyMetLysValSerSerLeuSerPheGlyAla 20
QY 70 TCTCCGCTCGGAAGTCTTCGGTCCAGTCCGCGAAGATGATCGCGCCACCGTGGC 129
DB 21 SerSerLeuGlyGlyValPheHisAspLeuLysGluGlyIleGlnAlaValPhe 40
QY 130 GAGGCTTTCCGCTCTCGGTATCAACTCTTCGACACCTCCCGGTATATGGAGGAACACTG 189
DB 41 ThrAlaIleGluAlaGlyMetAsnPheIleAspValSerProTyrTyrGlyHisTyrLys 60
QY 190 TCTGAGAAATGCTTGGTAAGGACTAAAGCTTTGCAAGTCCCTAGAGTGACTACATT 249
DB 61 AlaGluThrValLeuGlyLysAlaLeuLys-----AspIleProArgAspArgTyrTyr 78
QY 250 GTGGCTACTAGTGTGGTAGATAT---AAGNAGGT-----TTTGATTTCAGTGCT 297
DB 79 LeuSerThrLysValGlyArgTyrGlyLysAspGlyValAsnThrTyrAspTyrSerAla 98
QY 298 GAGAGGTAAGAAAGATTTGACGAGAGCTTGAGAGGCTTCAGCTTGATTATGTTGAC 357
DB 99 LysArgAlaThrGluSerValTyrGluSerMetGluArgLeuAsnIleAspPheIleAsp 118
QY 358 ATACTTCATGCCATGACATTCAGTTCGGGTCTCTTGATCATGATTCAGTGGAGAACAAATT 417
DB 119 LeuIleAsnValHisAspIleGluPheAlaAspLeuAsnGlnValValAsnGluThrLeu 138
QY 418 CTTGCTCTTCAGAACTGAAACAGAGGGGAGACCGGTTTCATTGGTATCATCTGCTT 477
DB 139 ProAlaLeuValGluLeuArgGluLysGlyValValGlyHisValGlyIleThrAspLeu 158
QY 478 CCGTTAGATATTTTCACTATGTTCTTCATCGAGTCCCTCCAGGACCTGTCGATGTGATA 537
DB 159 GlnLeuGluAsnLeuLysTyrPheValIleAspArgSerProSerGlyThrIleGluSerVal 178
QY 538 TTGTCTACTGTCAATACGGGCTTAATGATTCGAGCTTGGTGGATTTACTACCTTACTTG 597
DB 179 LeuSerPheCysHisTyrCysLeuLysAspAspLysLeuAlaAspPheLeuAspTyrPhe 198
QY 598 AAGACCAAGGTGGGTGTGATAGTCTTCTTCATTAGCATGGGCTCTTACAGAA 657
DB 199 GluSerLysGluIleGlyValIleAsnAlaSerProLeuSerMetGlyLeuSerGlu 218
QY 658 CAAGTCTCTCCTGAATGGCACCTCTCTCCCTGAGCTCAAGTCTGCAAGCAAGCCGCA 717
DB 219 ArgGlyValProValTyrPheIleAsnAlaProLysProLeuValAspAlaCysArgLysAla 238
QY 718 GTTGCTCACTGCAATCAAGGGCAAGAGATCACAAAGTTAGTCTCTGCAATACAGTTTA 777

DB 239 MetGluHisCysLysAlaLysAsnTyrProIleGluLysLeuAlaMetGlnPheSerVal 258
QY 778 GCAAACAAGAGATTTCGTCGGTTCGGATGAGCTCTGTCTCACAGGTAGAAGAA 837
DB 259 SerAsnProLysIleAlaThrThrLeuPheSerThrThrAsnProGluAsnValLysLys 278
QY 838 AATGTTGACAGTTTACAGAGCTTGAAGTCTGGGATGGATCAAGAACTCTGTCTGAG 897
DB 279 AsnIleGlyPheIleGluGlu-----ProValAspTyrGluLeuValArgGlu 294
QY 898 GTTGAAGCTATTCGAGCGCTGTAAAGATCTGACATGG 936
DB 295 ValArgGluIleIleGlyGluGlnArgValSerTyr 307

RESULT 6
Q7U240 PRELIMINARY; PRT; 347 AA.
AC Q7U240;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative oxidoreductase-possibly Aldo/keto reductase.
GN OrderedLocusNames=RB203;
OS Rhodopirellula Baltica.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416; DOI=10.1073/pnas.1431443100;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schleener H., Anann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303 (2003).
DR EMBL; BX294133; CAD71446.1; -.
DR InterPro; IPR001395; Aldo/ket_red.
DR Pfam; PF00248; Aldo_ket_red; 1.
DR ProDom; PD000288; Aldo/ket_red; 1.
KW Complete proteome.
SQ SEQUENCE 347 AA; 39075 MW; 4721A31577F8702C CRC64;

Alignment Scores:
Pred. No.: 4.45e-40 Length: 347
Score: 625.00 Matches: 123
Percent Similarity: 64.01% Conservative: 78
Best Local Similarity: 39.17% Mismatches: 103
Query Match: 37.81% Indels: 10
DB: 2 Gaps: 4

US-10-606-300-12 (1-960) x Q7U240 (1-347)
QY 19 CGAGCTTTGGGGAACACAGGCTTAAGTTAGCGCGTGGTGGTGGTCTCCGCTC 78
DB 36 ArgArgLeuGlyGlnThrAspMetGluLeuThrThrLeuSerPheGlyAlaSerSerIle 55
QY 79 GGAAGTGTCTTCGGTCCAGTCCCGGAGATATGCCGTGCCACCGTCCGCGAGGCTTTC 138
DB 56 GlyGlnGluPheArgSerValAspLeuGlyGluSerLeuGlnAlaValArgValAlaLeu 75
QY 139 CGTCTCGGTATCAACTCTTCGACACCTCCCGCTATTATTAGGAGGAACACTGTCTCAGAAA 198
DB 76 AspSerGlyMetAsnPheIleAspThrAlaAlaPheTyrGlyArgGlyMetSerGluMet 95
QY 199 ATGCTTGTGAAGGACTAAAGCTTTGCAAGTCCCTAGAGTGACTACATTGTGCTACT 258
DB 96 MetLeuGlyArgValLeuPro-----AspIleProArgAspGlnTyrTyrLeuGlyThr 113
QY 259 AAGTGTGTGATATATAA---GAAGTTTTCATTTTCAGTCTGAGAGAGTAAAGAGT 315
```

```
Db 114 LysLeuGlyArgTyThrGlyGluHisPheAspPheSerAlaLysArgValThrGluSer 133
QY 316 ATTGACGAGAGCTTGAGAGGCTTCAGCTTGATTTGTCATCATCTTCATTGCCATGAC 375
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 134 IleAspThrSerLeuGluArgMetLysValAspHisLeuAspIleValLeuCysHisAsp 153
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 376 ATTGAGTTCGGGTCTCTTGATCATGATGTGAGTGAACAATCTCTGCTTCAGAAATCG 435
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 154 LeuGluPheValGluMetSerGlnIleValGluGluThrIleProAlaIleArgArgGlu 173
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 436 AAACAGAGGGAAGACCGGTTCAATGTGATCACTGCTCTTCGTTAGATATTTTCACT 495
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 174 IleAlaLysGlyLysValArgTyThrValSerGlyTyThrProMetLysMetLys 193
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 496 TATGTTCTTGATCGAGTCCCGGACTCGCGACTGTCGATGTGATATTTGTCATCTGTCATTAC 555
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 194 TyrValMetAlaAsnThr-----AspIleAspCysLeuLeuThrTyrAsnHisTyr 210
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 556 GCGGTTAATGATTCGAGCTTGCTGCTGATTTACTACTTCTTGAAGCAAGAGTGTGGT 615
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 211 ThrLeuGlnAsnAspMetAlaLeuGluValProLeuAlaLysGluLysGlyValGly 230
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 616 GTGATAAGTCCTCTCCATTAGCAATGGCGCTCTTACAGAACAGGCTCCTCGAATGG 675
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 231 LeuMetAsnGlyAlaProPheSerAlaArgLeuLeuThrAsnAlaGluLeuProTyr 250
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 676 CACCCTGCTCCCTCGAGCTCAAGTCTGCAAGCAAGCGCGAGTGTCTCACTGCAAAATCA 735
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 251 HisLysAlaThrProGlnValArgGluValAlaAlaAlaAlaLysHisCysAlaAsp 270
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 736 AAGGCAAGAGATCACAAGTACTCTGCAATACAGTTAGCAAAAGAGAGATTCG 795
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 271 ArgGlySerAspIleAlaLysIleAlaLeuGlnPheSerIleAlaAsnGluAspPheAla 290
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 796 TCGGTGTTGGTGGGATGAGCTCTCTCACAGGTAGAGAAATGTTGCAGCAGTTACA 855
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 291 ThrCysIleProGlySerAlaAsnProAsnArgValAlaGlnTrpValGluTrpAlaGln 310
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 856 GAGCTTGAAAGTCTGGGATGATGATCAAGAAATCTGTCTGAGGTTGAAGCTATTCTCGAG 915
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 311 Glu-----ProIleAspGluThrLeuValAlaGluValLysGluIleLeuLys 326
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 916 CTGTGAAGAATCTGACATGGCCAGTGGATCCATCGAAC 957
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 327 ProfileHisAsnTrpPheTyrlleGluGlyLeuProGluAsn 340
RA ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
Q3VG3 ID Q9VGF3 PRELIMINARY; PRT; 345 AA.
AC Q9VGF3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CG18547-PA.
GN ORFName=CG18547;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.D.,
RA Abail J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballweir R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
```

```
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
RA Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jajali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maiti B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount D.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson K.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler P., Shen H.,
RA Shue B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassatman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RX FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RX FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003694; AAF54729.1; -.
DR HSP; P52895; 1J96.
DR FlyBase; FBgn0037973; CG18547.
```



```
Db 132 AspValValGlnIleHisAspValGluPheAlaProAsnLeuAspIleValLeuAsnGlu 151
QY 412 ACAATTCTGCTCTTCAGAAACTGAAACAAGAGGGGAAGACCCGTTTCATTGGTATCACT 471
Db 152 ThrLeuProThrLeuAlaLeuArgAlaGluGlyLeuAspGlyLeuValSer 171
QY 472 GGTCTTCGGTATAGATTTCCTTATGCTTCGTGAGTGCCTCCAGGAGCTGTCGAT 531
Db 172 AlaTyrProLeuAlaValLeuArgGluIleThrLeuAla---ProGlyArgPheAsp 190
QY 532 GTGATATTGTCATCTGTCATTCACGGCTTAATGATTCGAGCTTCTCGATTTACTACT 591
Db 191 ThrValLeuSerTyrCysArgAsnThrLeuPheAspSerLeuGluGluTyrIlePro 210
QY 592 TACTTGAAGAGCAAAAGGTGTGGTGTGATAGTGTCTTCCATTAGCAATGGCTCTCT 651
Db 211 PhePheArgGluHisLysMetGlyLeuIleCysAlaSerGlyHisGlyMetGlyLeuLeu 230
QY 652 ACAGAACAAAGTCTCTCGTAATGGCACCTGCTGCTCCCTGAGCTCAAGTCTGCAAGCAA 711
Db 231 ThrAsnGlyGlyProGlnProTyrHisProAlaAspArgGlnLeuArgGluValCysAla 250
QY 712 GCCGAGTGTCTCACTGCAATCAAGGGCAAGAGATCACAAATGATGCTGCAATAC 771
Db 251 GluAlaAlaIleTyrCysLysArgGluGlyValGluLeuGlyLysLeuAlaMetHisHis 270
QY 772 AGTTAGCAAAACAGGAGATTTCGTCGGTGTGGTGGATGAGCTGTCTCACAGGTA 831
Db 271 AlaLeuGlnMetProGlyProAlaThrPheLeuAlaGlyMetGlnThrProGluLeuVal 290
QY 832 GAAGAAATGTTGCACGAGTTACAGAGCTGAAAGCTGTGGGATGAGTCAAGAACTCTG 891
Db 291 GlnIleAsnLeuAlaTyrPhe-----GluGlyLeu 301
QY 892 TCTGAGTGTGAGCT---ATTCTCAGCTGTAAAGATCTGACATGGCCCA 939
Db 302 SerGluLysGluAlaAspValLeuSerTyrLeuLysGluArgValPhePro 318

RESULT 9
QSVGF1
ID QSVGF1 PRELIMINARY; PRT; 342 AA.
AC QSVGF1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE CG3397-PA (GH17980p).
CN ORFNames=CG3397;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.B.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
```

```
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang X., Lin X.,
RA Liu X., Mactai B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler P., Shen H.,
RA Shue B.C., Siden-Kimani I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RX FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBAJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RX FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBAJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBAJ databases.
```

[illegible]


```
Db 102 AlaAlaAlaIleGlyCysMetPheGlyAsnVal---GluAspSerIleIleLysIleVal 120
QY 127 CGCAGGCTTCCGCTCGTATCACTTCTTCGACACCTCCCGTATTATGAGGAAACA 186
Db 121 GluThrAlaIleLysGlnGlyLeuAspThrGlyTyrTrpTyrSerGlnSer 140
QY 187 CTGCTGAGAAATCTGGTGAAGGCACTAAAGGCTTTCGAAAGTCCCTAGAAAGTCACTAC 246
Db 141 ArgSerGluSerIleLeuGlyLysAlaLeuSer-----LysIleProArgLysAlaTyr 158
QY 247 ATTGGCTTACTAAGTGGGTAGA-----TATAAGAAGGTTTGTATTCAGT 294
Db 159 TyrIleSerThrLysValGlyArgPheGluLeuAspTyrAlaArgThrPheAspPheArg 178
QY 295 GCTGAGAGAGTAAAGAGTATTGACGAGCTTTCGAGAGCTTTCAGCTTCATGATGTT 354
Db 179 AlaAspLysIleLeuGluSerLeuThrAsnSerLeuLysArgGlnLeuThrTyrIle 198
QY 355 GACATA-----CTTCATTGCCATGACATTGAGTTCGGGTCTCTTTGATCAGATTGTG--- 405
Db 199 AspIleCysTyrValGlnIleHisAspAlaAspPheAlaProAsnGluSerIleValLeu 218
QY 406 AGTGAACAAATTCCTGCTCTTCAGAACTGAACAGAGGGGGAAGCCCGGTTCAATTGGT 465
Db 219 TyrGluThrLeuGlnAlaLeuGluMetAlaLysSerSerGlyLysIleArgHisIleGly 238
QY 466 ATCACTGGTCTCCGTTAGAT-----ATTTCACATGATGTTCTTCATCGAGTG 513
Db 239 LeuThrGlyTyrProLeuGlyLysLeuValTyrSerPheArgTyrValTrpAsn----- 256
QY 514 CTCCAGGAGCTGTCGATGTGATATGTCATACTGTCAATCGGCGTAAATGATTCGACG 573
Db 257 -----AlaProLysLeuTyrPheSerValAsnLeuSerIle 268
QY 574 TTGCTGGATTACTACCTTACTTGAAGAGCAAGGTGGGTGTCATAGTCTTCTCCA 633
Db 269 -----ThrArgAsnIleAlaValIleAsnSerGlyAla 279
QY 634 TTAGCAATGGGCTCTTACAGAACAAAGTCTCTCTGAATGGCACCCCTGCTTCCCTGAG 693
Db 280 LeuCysTrpGlyLeuLeuThrGluLysGlyProProTrpHisProAlaSerAspGlu 299
QY 694 CTCAGTCTGCAAGCAAGCCGAGTGTCTCACTGCAATCAAGGGCAAGAAAGATCACA 753
Db 300 IleLysGluAlaCysLeuAlaAlaThrThrTyrCysSerSerLysAsnIleSerIleSer 319
QY 754 AAGTATGCTCTGCAATACAGTTTACAAACAGGAGATTCCTCGGTGTTGGTTCGGATG 813
Db 320 LysLeuAlaLeuAspTyrAlaLeuAsnPheProAsnValIleCysCysLeuValGlyMet 339
QY 814 AGCTCTGCTCAAGTAGAAGAAATGTT-----GCAGCAGTTACAGAG 858
Db 340 AspSerValGlnGlnValLeuAspAsnLeuGluLeuSerAsnPheSerArgIleThrAsp 359
QY 859 CTTGAAAGTCTGGGAGTATGATCAAGAACTCTGTCTGAGGTTGAAGCTATTCTCGAGCT 918
Db 360 ValGluGlnArgValArgAspArg-----IleMetArgArgTyrLeuAspArg 375
QY 919 GTAAAGAACTGACATGG 936
Db 376 LeuGluAsnAlaGlyTrp 381
```

RESULT 11

Q8PF04

ID Q8PF04 PRELIMINARY; PRT; 336 AA.

AC Q8PF04;

DT 01-OCT-2002 (TrEMBLrel. 22, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Oxidoreductase.

GN OrderedLocusNames=XAC4184;

OS Xanthomonas axonopodis (pv. citri).

```
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
EX MEDLINE=2202145; PubMed=1204217; DOI=10.1038/417459a;
da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.P., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Teai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities";
RL Nature 417:459-463(2002).
DR EMBL; AE012069; AAM39019.1; -.
DR HSSP; P52895; IJ96.
DR InterPro; IPR001395; Aldo/ket red.
DR Pfam; PF00248; Aldo_ket_red; 1.
DR ProDom; PD000288; Aldo/Ket_red; 1.
KW Complete proteome.
SQ SEQUENCE 336 AA; 34982 MW; 8BF6EA56C8E68E97 CRC64;
```

Alignment Scores:

Pred. No.:	1.45e-24	Length:	336
Score:	426.00	Matches:	118
Percent Similarity:	50.86%	Conservative:	60
Best Local Similarity:	33.71%	Mismatches:	104
Query Match:	25.77%	Indels:	68
DB:	2	Gaps:	12

US-10-606-300-12 (1-960) x Q8PF04 (1-336)

```
QY 1 ATGACGAAATAGAGCTTCGAGCTTTGGGGAACACAGGGCTTAAGTTAGCGCCCTGGT 60
Db 1 MetSerGlyValAlaArgGlnArgLeuGlyGlyThrAspValGlnLeuSerAlaLeuGly 20
QY 61 TTTGGTGCTCTCCGCTCGGAGTGTCTTCGGTCCAGTCGCCGAGATGATGCGCTCGCC 120
Db 21 PheGlyAlaAlaProIleGlyAsnLeuTyrThrGluValAspGluAlaAspAlaLeuAla 40
QY 121 ACCGTGCGGAGGCTTTCCGTCTCGGTATCAACTTCTTCGACACCTCCCGCTATTATGA 180
Db 41 AlaValAlaGlyAlaPheGluAlaGlyIleArgHisPheAspThrAlaProTyrTyrGly 60
QY 181 GGAACATGCTCTGAGAAATGCTTGGTAAAGGCACTAAAGGCTTTCGAAAGTCCCTAGAGT 240
Db 61 TyrGlyLeuSerGluAlaArgLeuGlyArgGlyLeuAlaGly-----ValProArgAla 78
QY 241 GACTACATTTGGCTACTTAAGTGTGGTAGATAT----- 273
Db 79 AspTyrThrLeuSerThrLysValGlyArgCysValTyrAspAspAlaHisAlaAla 98
QY 274 ---AAAGAAGGT-----TTTGATTTTCAGTCTGAGAGA 303
Db 99 GlyArgAspGlyPheAlaValAlaGlyArgAlaGluPheAspTyrSerAlaAspGly 118
QY 304 GTAGAAAGAGTATTGACGAGAGCTTGGAGAGGCTTCAGCTTGATTATGTTGACATCTT 363
Db 119 ValArgArgAlaPheAlaSerSerLeuGlyArgLeuGlyThrAspTyrIleAspValLeu 138
QY 364 CATTGCCATGACATT-----GAGTTCGGGTCTCTTCATCAG 399
```


Db 139 LeuLeuHisAspIleGlyAlaLeuThrHisGlyAspAsnHisAlaAsnValLeuArgIn 158
QY 400 ATTGTGAGTGAACAATTCCTGCTCTCAGAACTGAACAAGGGGGAAGCCGGTTC 459
Db 159 AlalaLeuGluAlaLeuProAlaMetAlaGluLeuYsAlaAlaGlyAlaCysGlyAla 178
QY 460 ATTGGTATCAGTGGTCTTCGCTAGATATTTTCACTTATGTTCTTGTATCGATGCTCCCA 519
Db 179 IledGlyLeuGlyValAsnGluGlnAspValAlaLeuGluValLeuProArgPhePro-- 197
QY 520 GGGACTGCGATGATATTCATCTGCTCATTCAGCGGTATGATTCGACGTGCTG 579
Db 198 -----LeuAspCysValMetLeuAlaGlyArgTyr-----ThrLeuLeu 210
QY 580 GAT-----TTACTACCTTACTTGAAGAGCAAGGTGTGGTGTGATA 621
Db 211 GluGlnHisGlyAlaArgAlaLeuLeuGlnGlnAlaGlnArgAsnValAlaLeu 230
QY 622 AGTGTCTTCCATTGCAATGGCTCTCTACAGAA-----CAAGGTCTCTCTGAATGGCAC 678
Db 231 SerAlaGlyProTyrSerSerGlyLeuLeuSerAspAlaArgGlyProGlyAlaThrTyr 250
QY 679 CTGCTCTCCCT-----GAGCTCAAGTCTGCAAGCAAGCCGAGTCTCTCAC 726
Db 251 AsnTyrAlaProValAspThrAlaThrLeuGlnHisAlaGlnArgLeuTyrAlaAla--- 269
QY 727 TGCAAATCAAGGGCAAGATCAAAAGTTAGCTCTGCAATCAGTTTACGAACAAG 786
Db 270 CysAlaAlaPheGlyValAspIleGlyAlaAlaLeuGlnPheProLeuAlaHisPro 289
QY 787 GAGATTTCGTGCTGTTGTTGGGATGAGTCTGCTCACAGGTAGAGAAATGTTGCA 846
Db 290 AlaValThrValValAlaGlyMetArgSerValAlaGluValArg-----Ser 306
QY 847 GCAGTTCAGAGCTTGAAGTCTGGGATGGATGATCAAGAACTCTCTGAGGTGAAGCT 906
Db 307 AlaAlaThrArgLeuAlaAla----- 313
QY 907 ATTCTCGAGCTGTGAAGATCTGACATGG 936
Db 314 -----ProValProSerAlaLeuTyr 320
RESULT 12
ID Q9VGF2 PRELIMINARY; PRT; 294 AA.
AC Q9VGF2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE CG12224-PA.
GN ORFNames=CG12224;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.D.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Paclos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattle B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacie J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacie J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirekas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB003694; AAF54730.1; -;
DR FlyBase; FBgn0037974; CG12224.
DR InterPro; IPR001395; Aldo/ket_red.
DR Pfam; PF00248; Aldo_ket_red; 1.
DR PRINTS; PR00069; ALDETTRDTASE.
SQ SEQUENCE 294 AA; 32984 MW; 8B03E0F0C8EE1B8E CRC64;

```
Alignment Scores:
Pred. No.: 3,78e-24 Length: 294
Score: 420.50 Matches: 102
Percent Similarity: 52.48% Conservative: 57
Best Local Similarity: 33.66% Mismatches: 101
Query Match: 25.44% Indels: 43
Db: 2 Gaps: 7

US-10-606-300-12 (1-960) x Q9VGF2 (1-294)
QY 7 AATATAGAGCTTCGAGCTTTGGGGAACACAGGCTTTAAAGTTAGCGCGTGGTTTGGT 66
Db 7 AATATAGAGCTTCGAGCTTTGGGGAACACAGGCTTTAAAGTTAGCGCGTGGTTTGGT 66
QY 21 ArgMetGluTyrArgGlnLeuHisValSerLeuAlaLeuGly 40
Db 21 ArgMetGluTyrArgGlnLeuHisValSerLeuAlaLeuGly 40
QY 67 GCTCTCCGCTCGGAAGTGTCTTCGCTCCAGTCGCGAAGATGATGCGCGCCACCGTG 126
Db 67 GCTCTCCGCTCGGAAGTGTCTTCGCTCCAGTCGCGAAGATGATGCGCGCCACCGTG 126
QY 41 GlySerProLeuCyAsnLeu----- 47
Db 41 GlySerProLeuCyAsnLeu----- 47
QY 127 CGCAGGCTTCCGCTCGGTATCACTTCTTCGACACCTCCCGTATTTAGGAGAAC 186
Db 127 CGCAGGCTTCCGCTCGGTATCACTTCTTCGACACCTCCCGTATTTAGGAGAAC 186
QY 48 -----PhePheAspTyrAspArgGluGlyLe 58
Db 48 -----PhePheAspTyrAspArgGluGlyLe 58
QY 187 CTGCTCTGAGAAATGCTTGGTAAAGGACTTAAAGGCTTTGCAAGTCCCTAGAGTCACTAC 246
Db 187 CTGCTCTGAGAAATGCTTGGTAAAGGACTTAAAGGCTTTGCAAGTCCCTAGAGTCACTAC 246
QY 59 Leu-----MetAlaLeuLysAspValProArgGluAlaTyr 70
Db 59 Leu-----MetAlaLeuLysAspValProArgGluAlaTyr 70
QY 247 ATTGTGGCTACTAAGTGTGGTAGATAT-----AAAGAAGGTTTGAATTCAGT 294
Db 247 ATTGTGGCTACTAAGTGTGGTAGATAT-----AAAGAAGGTTTGAATTCAGT 294
QY 71 TyrIleAlaThrLysValAlaArgTyrGlyLeuAspProLysAsnMetPheAspTyrSer 90
Db 71 TyrIleAlaThrLysValAlaArgTyrGlyLeuAspProLysAsnMetPheAspTyrSer 90
QY 295 GCTGAGAGTAAAGAGTATTACGAGACCTTGGAGAGCTTCAGCTTGATTTATGTT 354
Db 295 GCTGAGAGTAAAGAGTATTACGAGACCTTGGAGAGCTTCAGCTTGATTTATGTT 354
QY 91 AlaAspLysAlaSerValLysArgSerLeuGluArgLeuAspArgVal 110
Db 91 AlaAspLysAlaSerValLysArgSerLeuGluArgLeuAspArgVal 110
QY 355 GACATCTTCATTCGATGACATTCAGTTTCGGG---TCTCTTGATCAGATTCAGTCAA 411
Db 355 GACATCTTCATTCGATGACATTCAGTTTCGGG---TCTCTTGATCAGATTCAGTCAA 411
QY 111 AspIleLeuGlnValHisAspValAspAlaProAsnLeuAspIleValLeuAsnGlu 130
Db 111 AspIleLeuGlnValHisAspValAspAlaProAsnLeuAspIleValLeuAsnGlu 130
QY 412 ACAATTCCTGCTCTTCAGAACTGAACAGAGGGGAGACCCGTTTCATTTGATCATCT 471
Db 412 ACAATTCCTGCTCTTCAGAACTGAACAGAGGGGAGACCCGTTTCATTTGATCATCT 471
QY 131 ThrIleProValLeuGluGlyValGlnAlaGlyLysAlaArgPheIleGlyValThr 150
Db 131 ThrIleProValLeuGluGlyValGlnAlaGlyLysAlaArgPheIleGlyValThr 150
QY 472 GCTCTCCGTTAGATATTTCTATGTTCTTGTGATCGAGTCCCTCCAGGACTTCGAT 531
Db 472 GCTCTCCGTTAGATATTTCTATGTTCTTGTGATCGAGTCCCTCCAGGACTTCGAT 531
QY 151 AlaTyrAspValAspValLeuLysGluCysAlaGluArg---GlyLysGlyArgIleGln 169
Db 151 AlaTyrAspValAspValLeuLysGluCysAlaGluArg---GlyLysGlyArgIleGln 169
QY 532 GTGATATTGTCATCTGTCATTCACGCGTTAATGATTCGACGTTGCTGATTTACTACT 591
Db 532 GTGATATTGTCATCTGTCATTCACGCGTTAATGATTCGACGTTGCTGATTTACTACT 591
QY 170 ValValLeuAsnTyrAlaArgTyrThrLeuLeuAspAsnThrLeuLeuArgTyrMetLys 189
Db 170 ValValLeuAsnTyrAlaArgTyrThrLeuLeuAspAsnThrLeuLeuArgTyrMetLys 189
QY 592 TACTTGAAGACAAAGGTGGTGTGATAAGTGTCTTCCATTAGCAATGGGCTTCCTT 651
Db 592 TACTTGAAGACAAAGGTGGTGTGATAAGTGTCTTCCATTAGCAATGGGCTTCCTT 651
QY 190 AspPheGlnLysMetGlyValGlyValValCysAlaAlaHisSerLeuGlyLeuLeu 209
Db 190 AspPheGlnLysMetGlyValGlyValValCysAlaAlaHisSerLeuGlyLeuLeu 209
QY 652 ACAGAACAGGCTCTCTCAATGGGACCTGCTTCCCTGAGCTCAAGTCTCGACAGCAA 711
Db 652 ACAGAACAGGCTCTCTCAATGGGACCTGCTTCCCTGAGCTCAAGTCTCGACAGCAA 711
QY 210 ArgAsnAlaGlyProHisAlaSerHisProGlySerGlnGluIleAlaValAlaLys 229
Db 210 ArgAsnAlaGlyProHisAlaSerHisProGlySerGlnGluIleAlaValAlaLys 229
QY 712 GCCGAGTTGCTCACTGCAATCAAGGGCAAGAGATCACAAGGTAGCTCTGCAATAC 771
Db 712 GCCGAGTTGCTCACTGCAATCAAGGGCAAGAGATCACAAGGTAGCTCTGCAATAC 771
QY 230 ArgGlyAlaGluIleCysGlnGlnArgAsnValGluLeuGlyLysLeuAlaMetTyrTyr 249
Db 230 ArgGlyAlaGluIleCysGlnGlnArgAsnValGluLeuGlyLysLeuAlaMetTyrTyr 249
QY 772 AGTTTAGCAACAGAGATTTGCTGGTGGTGGTGGATGAGCTGCTCTCAAGGTA 831
Db 772 AGTTTAGCAACAGAGATTTGCTGGTGGTGGTGGATGAGCTGCTCTCAAGGTA 831
QY 250 ThrMetGlnLeuAspGlyAlaAlaThrPheLeuIleGlyIleProAsnArgLysLeuLeu 269
Db 250 ThrMetGlnLeuAspGlyAlaAlaThrPheLeuIleGlyIleProAsnArgLysLeuLeu 269
QY 832 GAAGAAATGTCGACGCTTACAGACTTGAAGTCTGGGATG-----GATCAA 882
Db 832 GAAGAAATGTCGACGCTTACAGACTTGAAGTCTGGGATG-----GATCAA 882
QY 270 ArgIleAsnLeuAspAlaIlePheAsp-----GlyLeuThrSerHisGluGln 285
Db 270 ArgIleAsnLeuAspAlaIlePheAsp-----GlyLeuThrSerHisGluGln 285
QY 883 GAAACTCTG 891
Db 883 GAAACTCTG 891
QY 286 GluValLeu 288
Db 286 GluValLeu 288

RESULT 13
Q8P3K6 PRELIMINARY; PRT; 335 AA.
AC Q8P3K6;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Oxidoreductase.
GN OrderedLocusNames=XCC4065;
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=2202145; PubMed=12024217; DOI=10.1038/417459a;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Parah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
RA Cicarelli R.M.B., Cannavan F., Cardoso J., Chamberg F., Ciapina L.P.,
RA Paria J.B., Ferreira A.J.S., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Fornighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL: AE012528; AAM43286.1;
DR InterPro: IPR001395; Aldo/ket_red.
DR Pfam: PF00248; Aldo_ket_red; 1.
DR ProDom: PD000288; Aldo/ket_red; 1.
KW Complete proteome.
SQ SEQUENCE 335 AA; 34773 MW; 36B212C16AFA2F1F CRC64;

Alignment Scores:
Pred. No.: 6.65e-24 Length: 335
Score: 417.50 Matches: 108
Percent Similarity: 52.77% Conservative: 54
Best Local Similarity: 35.18% Mismatches: 98
Query Match: 25.26% Indels: 47
Db: 2 Gaps: 9

US-10-606-300-12 (1-960) x Q8P3K6 (1-335)
QY 25 TTGGGACACACAGGCTTAAAGTTAGCGCGTGGTTTGGTGGCTTCGCTCGGAAAGT 84
Db 25 TTGGGACACACAGGCTTAAAGTTAGCGCGTGGTTTGGTGGCTTCGCTCGGAAAGT 84
QY 8 ValGlySerThrAlaValGlnLeuSerAlaLeuGlyPheGlyAlaAlaProIleGlyAsn 27
Db 8 ValGlySerThrAlaValGlnLeuSerAlaLeuGlyPheGlyAlaAlaProIleGlyAsn 27
QY 85 GTCTTCGCTCGAGTCGCGCAAGATGATCCGTCGCCACCGTCGCGGAGGCTTTCGCTC 144
Db 85 GTCTTCGCTCGAGTCGCGCAAGATGATCCGTCGCCACCGTCGCGGAGGCTTTCGCTC 144
QY 28 LeuTyrSerAlaValAspGluGlyAspAlaLeuAlaValThrGlyAlaPheGluAla 47
Db 28 LeuTyrSerAlaValAspGluGlyAspAlaLeuAlaValThrGlyAlaPheGluAla 47
QY 145 GGTATCACTTCTTCGACACCTCCCGCTATTATGGAGGACACACTCTCTGAGAAATGCTT 204
Db 145 GGTATCACTTCTTCGACACCTCCCGCTATTATGGAGGACACACTCTCTGAGAAATGCTT 204
QY 48 GlyIleArgHisPheAspAlaAlaProTyrTyrGlyTyrGlyLeuSerGluAlaArgLeu 67
Db 48 GlyIleArgHisPheAspAlaAlaProTyrTyrGlyTyrGlyLeuSerGluAlaArgLeu 67
QY 205 GGTAGGACTAAAGGCTTTCAGACTCCCTAGAGTACTACATTGGTGGCTACTAAGTGT 264
Db 205 GGTAGGACTAAAGGCTTTCAGACTCCCTAGAGTACTACATTGGTGGCTACTAAGTGT 264
QY 68 GlyArgGlyLeuAlaGly-----ValProArgAlaAspTyrThrLeuSerThrLysVal 85
Db 68 GlyArgGlyLeuAlaGly-----ValProArgAlaAspTyrThrLeuSerThrLysVal 85
QY 265 GGTAGATAT-----AAAGAAGT----- 282
Db 265 GGTAGATAT-----AAAGAAGT----- 282
QY 86 GlyArgCysValTyrAspAlaThrAlaValAlaGlyArgAspGlyPheAlaValAla 105
Db 86 GlyArgCysValTyrAspAlaThrAlaValAlaGlyArgAspGlyPheAlaValAla 105
```


QY 79 GGAAGTGTCTTCGGTCCAGTCGCCGAAGATGAT-----GCCGTCGCCACC 123

Search completed: November 13, 2005, 08:30:17
Job time : 122.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 13, 2005, 08:37:08 ; Search time 614 Seconds
(without alignments)

3075.566 Million cell updates/sec

Title: US-10-606-300-11

Perfect score: 1619

Sequence: 1 MTKIELRALGNTGLKVSAGV.....VEAILEPQNLTPWPSGIHQN 319

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=xlh
-Q=/Cgn2_1/USPTO-epool_h/US10606300/runat_07112005_094937_8159/app_query.fasta_1.455
-DB=N Geneseq -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bloum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HRAPEXT=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10606300@cgn_1_1_470 @runat_07112005_094937_8159 -NCPU=6 -ICPU=3
-NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq_16Dec04:*

1: geneseqn1980s:*

2: geneseqn1990s:*

3: geneseqn2000s:*

4: geneseqn2001as:*

5: geneseqn2001bs:*

6: geneseqn2002as:*

7: geneseqn2002bs:*

8: geneseqn2003as:*

9: geneseqn2003bs:*

10: geneseqn2003cs:*

11: geneseqn2003ds:*

12: geneseqn2004as:*

13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1619	100.0	960	AAD19526	Arabidopsis
2	1619	100.0	960	ABK10125	A. thaliana
3	1588	98.1	1221	AAAC37387	Arabidopsis
4	779.5	48.1	564	ACN48774	Cotton pr
5	593.5	36.7	1038	ABL20521	Drosophila

c	6	546.5	33.8	3724	4	ABL20520	Drosophila
c	7	544.5	33.6	1051	4	ABL28825	Drosophila
c	8	517	31.9	3308	4	ABL28824	Drosophila
c	9	497	30.7	563	13	ACN48848	Cotton pr
c	10	458	28.3	3014	4	ABL06602	Drosophila
c	11	420.5	26.0	885	4	ABL06603	Drosophila
c	12	366	22.6	987	2	AAQ62089	L-fucose
c	13	342	21.1	10400	4	ABL20064	Drosophila
c	14	342	21.1	10418	4	ABL28814	Drosophila
c	15	342	21.1	10451	4	ABL19246	Drosophila
c	16	339.5	21.0	2679	2	AAQ28895	Fucose de
c	17	337.5	20.8	921	13	ADT48436	Bacterial
c	18	329.5	20.4	912	8	ACA21511	Prokaryot
c	19	327	20.2	975	13	ADS50132	Bacterial
c	20	322	19.9	924	13	ADS55648	Bacterial
c	21	322	19.9	987	8	ACA33777	Prokaryot
c	22	317	19.6	1104	11	ABD11443	Pseudomon
c	23	317	19.6	2514	11	ABD11187	Pseudomon
c	24	316	19.5	921	13	ADT41709	Bacterial
c	25	316	19.5	1038	8	ACA31676	Prokaryot
c	26	314	19.4	349980	6	ABQ81847	Bifidobac
c	27	313	19.3	981	4	AA52452	E. coli D
c	28	313	19.3	981	8	ACA19190	Prokaryot
c	29	313	19.3	981	13	ADS46119	Bacterial
c	30	311	19.2	1275	6	ABA96767	Coriolus
c	31	310.5	19.2	906	8	ACA47550	Prokaryot
c	32	310	19.1	933	8	ACA21491	Prokaryot
c	33	307.5	19.0	933	13	ADT44628	Bacterial
c	34	306.5	18.9	999	8	ACA50894	Prokaryot
c	35	305.5	18.9	972	4	AA54112	Pseudomon
c	36	302.5	18.9	999	8	ACA51519	Prokaryot
c	37	302.5	18.7	1026	11	ABD14901	Pseudomon
c	38	302.5	18.7	1515	11	ABD15344	Pseudomon
c	39	302.5	18.7	20000	9	ADA00836	Agrobacte
c	40	301.5	18.6	1017	13	ADT45414	Bacterial
c	41	300.5	18.6	960	13	ADT44646	Bacterial
c	42	297.5	18.4	936	6	ABK72756	Bacillus
c	43	296.5	18.3	945	8	ACA33942	Prokaryot
c	44	296.5	18.3	999	10	ADC3610	E. faeciu
c	45	296	18.3	1005	9	ADA30666	DNA encod

ALIGNMENTS

RESULT 1
AAD19526
ID AAD19526 standard; CDNA; 960 BP.
XX AC AAD19526;
XX DT 18-DEC-2001 (first entry)
XX DE Arabidopsis thaliana L-galactose dehydrogenase (L-galdH) cDNA.
XX KW L-galactose dehydrogenase; L-galdH; transgenic organism; ascorbic acid;
XX KW herbicide; chromosome 4; ss.
XX OS Arabidopsis thaliana.
XX FH Key Location/Qualifiers
XX FT CDS 1..960
XX FT /*tag= a
XX FT /product= "L-galactose dehydrogenase (L-galdH)"
XX PN WO200172974-A2.
XX PD 04-OCT-2001.
XX PF 29-MAR-2001; 2001WO-GB001412.
XX PR 29-MAR-2000; 2000GB-00007651.
XX PA (ASCO-) ASCORBEX LTD.

```
XX Smirnoff N, Wheeler G;
XX
XX WPI; 2001-616482/71.
XX P-PSDB; AAE11998.
XX Novel L-galactose dehydrogenase protein and nucleic acid sequence
PT encoding the protein for producing genetically modified plants and
PT microorganisms with enhanced ability to synthesize ascorbic acid.
XX
XX Claim 6; Page 56-57; 58pp; English.
XX
XX The present invention relates to an isolated protein having L-galactose
CC dehydrogenase (L-galdH) biological activity. L-galdH nucleic acid is
CC useful for generating transgenic organisms and modified plants with
CC enhanced ability to synthesize ascorbic acid. L-galdH facilitate the
CC production of a plant that has been genetically modified to express a
CC mutated L-galdH protein which is resistant to herbicides that act against
CC the naturally occurring L-galdH and to identify and/or design compounds
CC that are inhibitors of L-galdH. The compounds can be used, for e.g. in a
CC herbicide which acts on L-galdH and damages or kills plants that express
CC the enzyme. The present sequence is Arabidopsis thaliana L-galactose
CC dehydrogenase (L-galdH) cDNA. The L-galdH gene is located on chromosome 4
XX
XX Sequence 960 BP; 251 A; 195 C; 251 G; 263 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 3,46e-163 Length: 960
Score: 1619.00 Matches: 319
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0
US-10-606-300-11 (1-319) x AAD19526 (1-960)
QY 1 MetThrLysIleGluLeuArgAlaLeuGlyAenThrGlyLeuLysValSerAlaValGly 20
Db 1 ATGACGAAATAGAGCTTCGAGCTTGGGGAACACAGGCGCTTAAGGTTAGCGCGCTTGGT 60
QY 21 PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAla 40
Db 61 TTGTGTGCTCTCCGCTCGAAGATGCTTCGGTCCAGTCGCGAAGATGATCGCGTCGCC 120
QY 41 ThrValArgGluAlaPheArgLeuGlyIleAenPheAspThrSerProTyrTyrGly 60
Db 121 ACCGTGCGGAGGCTTTCGGTCTCGGTATCACTTCTCGACACCTCCCGTATTATGGA 180
QY 61 GlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSer 80
Db 181 GGAACACTGCTCTGAGAAATGCTTGGTAAGGGACTTAAAGGCTTTTGCAGAGTCCCTAGAAGT 240
QY 81 AspTyrIleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGlu 100
Db 241 GACTACATTGTGGCTACTAAGTGTGGTATATTAAGAGGTTTGTATTTAGTCTGAG 300
QY 101 ArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 120
Db 301 AGAGTAAGAAAGAGTATTGACGAGAGCTTGGAGAGCTTCAGCTTGATATTGTTGACATA 360
QY 121 LeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePro 140
Db 361 CTTTCATTGTCATGACATTGAGTTCGGTCTCTTGATCAGATTGTGAGTCAAAACAAATTCCT 420
QY 141 AlaLeuGlnLysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGlyLeuPro 160
Db 421 GCTCTTCAGAAACTGAACAGAGGGGAGACCCGGTTCATTGGTATCAGTGTCTTCG 480
QY 161 LeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeu 180
Db 481 TTAGATATTTCACCTATGTTCTTCATCGAGTGCCTCCAGGAGCTGTCGATGTGATATTG 540
QY 181 SerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLys 200
```

```
Db 541 TCATACTGTCTATTACGGCGTTAATGATTGCGAGTGTGCTGGATTTTACTACCTTACTTGAAG 600
QY 201 SerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGln 220
Db 601 AGCAAGGTGTGGGTGTGATAGTCTTCTCATTTAGCAATGGGCTCTCTTACAGAACAA 660
QY 221 GlyProGluTyrHisProAlaSerProGluLeuLysSerAlaSerLysAlaVal 240
Db 661 GGTCTCTCTGAATGGCACCCCTGCTCCCTGAGCTCAAGTCTGCAAGCAAGCGCAGTT 720
QY 241 AlaHisCysLysSerLysGlyLysIleThrLysLeuAlaLeuGlnTyrSerLeuAla 260
Db 721 GCTCACTCAATCAAGGCGCAAGATCAAAAGTTAGCTCTGCAATACAGATTAGCA 780
QY 261 AsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGluAen 280
Db 781 AACAGGAGATTTCTGCTGGTGTGGTGGATGAGCTCTGCTCACAGTTAGAGAAAT 840
QY 281 ValAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluVal 300
Db 841 GTTGACGAGTTACAGAGCTTGAAGTCTGGGATGGATCAAGAACTCTGCTGAGGTT 900
QY 301 GluAlaLeuLeuGluProValLysAsnLeuThrTyrProSerGlyIleHisGlnAsn 319
Db 901 GAAGCTATTCTCGAGCCTGTAAGAATCTGACATGGCCCAAGTGAATCCATCAGAAC 957
RESULT 2
ABK10125
ID ABK10125 standard; cDNA; 960 BP.
XX
AC ABK10125;
XX
DT 21-MAY-2002 (first entry)
XX
DE A. thaliana cDNA encoding L-galactose dehydrogenase (LGDH).
XX
KW ss; gene; ascorbic acid; vitamin C; scurvy; recombinant yeast;
KW L-galactose dehydrogenase; LGDH; L-galactono-1,4-lactone dehydrogenase;
KW AGD; D-arabino dehydrogenase; ARA; D-arabinono-1,4-lactone oxidase;
KW ALO; L-gulono-1,4-lactone oxidase; GLO; aldonolactonase; AL.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 1..960
FT /tag= a
FT /product= "LGDH"
XX
XX WO200210425-A2.
XX
XX 07-FEB-2002.
XX
XX 02-AUG-2001; 2001WO-GB003485.
XX
XX 02-AUG-2000; 2000US-00630983.
XX
XX (BIOP-) BIOPOL SCARL.
XX (WHAL/) WHALLEY K.
XX
XX Porro D, Sauer M;
XX
XX WPI; 2002-217125/27.
XX P-PSDB; AAU76343.
XX
XX Generating ascorbic acid or its salt, involves culturing yeast capable of
XX converting ascorbic acid precursor into ascorbic acid in medium
XX comprising ascorbic acid precursor, and isolating ascorbic acid.
XX
XX Claim 14; Page 87; 95pp; English.
XX
XX The invention relates to generating ascorbic acid or its salt, involves
XX obtaining a recombinant yeast capable of converting an ascorbic acid
```

CC precursor into ascorbic acid, culturing the recombinant yeast in a medium
 CC comprising an ascorbic acid precursor, thus forming ascorbic acid, and
 CC isolating the ascorbic acid. Also include are stabilising ascorbic acid
 CC or its salt in a medium, by culturing a yeast in a medium comprising
 CC ascorbic acid or its salt and a recombinant yeast functionally
 CC transformed with a coding region encoding a protein having an enzyme
 CC activity selected from L-galactose dehydrogenase (LGDH), L-galactono-1, 4
 CC -lactone dehydrogenase (AGD), D-arabinose dehydrogenase (ARA), D-
 CC arabinono-1,4-lactone oxidase (ALO), L-gulonono-1,4-lactone oxidase (GLO)
 CC and aldono-lactonase (AL) activity, where the recombinant yeast is capable
 CC of converting to ascorbic acid at least about 25% ascorbic acid precursor
 CC or is capable of producing at least 20 mg ascorbic acid/L medium, when
 CC the yeast is cultured in a medium comprising one ascorbic acid precursor.
 CC The ascorbic acid produced (Vitamin C) is a powerful antioxidant, a
 CC deficiency of which causes scurvy in humans. The present sequence encodes
 CC A. thaliana LDGH

XX
 SQ Sequence 960 BP; 251 A; 195 C; 251 G; 263 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3.46e-163 Length: 960
 Score: 1619.00 Matches: 319
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-606-300-11 (1-319) x ABK10125 (1-960)

Qy 1 MetThrIysIleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGly 20
 Db 1 ATGACGAAATAGAGCTTCGAGCTTGGGGAACACAGGCTTAAAGGTTAGCGCGTTGGT 60
 Qy 21 PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAla 40
 Db 61 TTTGGTCCCTCTCGCTCGGAAGTGTCTTCGGTCCAGTCCGCGAAGATGATGCCGTGCC 120
 Qy 41 ThrValArgGluAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyrTyrGly 60
 Db 121 ACCGTGCGGAGGCTTTCGGTCTCGGTATCACTTCTCGACACTCCCGTATTATGGA 180
 Qy 61 GlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSer 80
 Db 181 GGAACACTGTCTGAGAAATGCTTGGTAAGGACTTAAAGGCTTTCGAAAGTCCCTAGAGT 240
 Qy 81 AspTyrIleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGlu 100
 Db 241 GACTACATTTGGCTACTAAGTGTGTAGATATATAAGAAAGGTTTGTATTCAGTGTGAG 300
 Qy 101 ArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 120
 Db 301 AGAGTAAAGAGATATTGACGAGAGCTTGGAGAGGCTTCAGCTTGATATGTTGACATA 360
 Qy 121 LeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePro 140
 Db 361 CTTCAATGCCATGACATTGAGTTCCGGTCTCTTGATCAGATTGTGAGTGAACAATTCCT 420
 Qy 141 AlaLeuGlnLysLeuLysGluGluGlyThrArgPheIleGlyIleThrGlyLeuPro 160
 Db 421 GCTCTTCAGAACTGAAACAAGAGGGGAAGACCCGGTTTCATTTGGTATCAGTGTCTCCG 480
 Qy 161 LeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeu 180
 Db 481 TTAGATATTTTCACTTATGTTCTTGATCGAGTGCTCCAGGAGCTGTGATGTGATTTG 540
 Qy 181 SerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLys 200
 Db 541 TCATACTGTCTATTACGGGTTAAATGATTCGAGCTTGTGGATTTACTACCTTACTTGAAG 600
 Qy 201 SerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGln 220
 Db 601 AGCAAGGTGTGGGTGTGATAGTGTCTTCCATTAGCAATGGGCTCTTACAGAACAA 660

Qy 221 GlyProProGluTrrHisProAlaSerProGluLeuLysSerAlaSerLysAlaVal 240
 Db 661 GGTCTCTCTGAATGGCACCCCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAGCCGAGTT 720
 Qy 241 AlaHisCysLysSerLysGlyLysLeuIleThrLysLeuAlaLeuGlnTyrSerLeuAla 260
 Db 721 GCTCACTGCAAAATCAAGGGCAAGAGATCAAAAGTTAGCTCTGCAATACAGTTTACGA 780
 Qy 261 AsnLysGluLysSerSerValLeuValGlyMetSerValSerGlnValGluGluAsn 280
 Db 781 AACCAAGGAGATTTCTGTCGGTGTGTTGGGATGAGCTCTGTCTCACAGGTAGAGAAAT 840
 Qy 281 ValAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluVal 300
 Db 841 GTTCAGCAGTTACAGAGCTTGAAGTCTGGGGATGGATCAAGAACTCTCTCTGAGGTT 900
 Qy 301 GluAlaIleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGlnAsn 319
 Db 901 GAAGCTATTTCTGAGGCTGTAAAGANCTGACATGGCCAGTGGANCTCATCAGNAC 957

RESULT 3
 AAC37387
 ID AAC37387 standard; DNA; 1221 BP.
 AC AAC37387;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 17196.
 XX
 KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway; metabolic pathway;
 KW promoter; termination sequence; ss.
 XX
 OS Arabidopsis thaliana.
 XX
 FN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-00301439.
 XX
 PR 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 09-MAR-1999; 99US-0123548P.
 PR 23-MAR-1999; 99US-0125788P.
 PR 25-MAR-1999; 99US-0126284P.
 PR 29-MAR-1999; 99US-0126785P.
 PR 01-APR-1999; 99US-0127462P.
 PR 06-APR-1999; 99US-0128234P.
 PR 08-APR-1999; 99US-0128714P.
 PR 16-APR-1999; 99US-0129845P.
 PR 19-APR-1999; 99US-0130077P.
 PR 21-APR-1999; 99US-0130449P.
 PR 23-APR-1999; 99US-0130510P.
 PR 28-APR-1999; 99US-0130891P.
 PR 30-APR-1999; 99US-0131449P.
 PR 30-APR-1999; 99US-0132048P.
 PR 30-APR-1999; 99US-0132407P.
 PR 04-MAY-1999; 99US-0132484P.
 PR 05-MAY-1999; 99US-0132485P.
 PR 06-MAY-1999; 99US-0132486P.
 PR 06-MAY-1999; 99US-0132487P.
 PR 07-MAY-1999; 99US-0132863P.
 PR 11-MAY-1999; 99US-0134256P.
 PR 14-MAY-1999; 99US-0134218P.
 PR 14-MAY-1999; 99US-0134219P.
 PR 14-MAY-1999; 99US-0134221P.
 PR 14-MAY-1999; 99US-0134370P.
 PR 18-MAY-1999; 99US-0134768P.
 PR 19-MAY-1999; 99US-0134941P.
 PR 20-MAY-1999; 99US-0135124P.
 PR 21-MAY-1999; 99US-0135353P.

PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.

PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149436P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 22-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.

PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:

Pred. No.: 1,02e-159 Length: 1221
Score: 1588.00 Matches: 317
Percent Similarity: 99.69% Conservative: 1
Best Local Similarity: 99.37% Mismatches: 1
Query Match: 98.09% Indels: 1
DB: 3 Gaps: 0

US-10-606-300-11 (1-319) x AAC37387 (1-1221)

Qy 1 MetThrLysIleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGly 20
Db 116 ATGACGAAATAGAGCTTCA-GCTTTGGGGAACACAGGCTTAAGTTTACCGCGTGGT 174
Qy 21 PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAla 40
Db 175 TTTGGTGCTCTCCGCTCGGAAGTGCTCTCGGTCCAGTCGCGGAAGATGATCGCGTCGCC 234
Qy 41 ThrValArgGluAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyrTyrGly 60
Db 235 ACCGTGCGGAGGCTTTCCGTCTCGGTATCACTTCCTTCGACACCTCCCGTATTATGGA 294
Qy 61 GlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSer 80
Db 295 GGAACACTGTCTGAGAAATGCTTGGTAAGGACTAAGGCTTTCGAAGTCCCTAGAGT 354
Qy 81 AspTyrIleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGlu 100
Db 355 GACTACATTGTGGCTACTAAGTGTGTAGATATAAAGAAGGTTTTGATTTTCAGTGTGAG 414
Qy 101 ArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 120
Db 415 AGAGTAGAAAGAGATTGACGAGAGCTTGAGAGGCTTCAGCTTGATATTATGTTGACATA 474
Qy 121 LeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePro 140
Db 475 CTTTATTTGCCATGACATTGAGTTGGGTCTCTTATCATGATGTGAGTGAACATTCCT 534
Qy 141 AlaLeuGlnLysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGlyLeuPro 160
Db 535 GCTCTTCAGAACTGAAACAGAGGGGAGACCCGGTTCAATTGTATGCTACCTGCTCTCCG 594
Qy 161 LeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeu 180
Db 595 TTAGATATTTTCACTTATGTTCTTGATCGAGTGCTCCAGGAGCTGTGATGTGATATTG 654
Qy 181 SerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLys 200
Db 655 TCATCTGTCTATTACGGGTAAATGATTCGAGCTTGCTGGATTTACTACCTTACTTGAAG 714
Qy 201 SerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGln 220
Db 715 AGCAAAAGTGGGTGTGATAAGTGCTTCTCCATTAGCAATGGGCTCTTACAGAACAA 774
Qy 221 GlyProProGluThrHisProAlaSerProGluLeuLysSerAlaSerLysAlaVal 240
Db 775 GGTCTCTCGAATGCGACCCCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAGCGCGCATTT 834
Qy 241 AlaHisCysLysSerLysGlyLysValIleThrLysLeuAlaLeuGlnTyrSerLeuAla 260
Db 835 GCTCACTCAATCAAGGGCAAGAGATCAAAAGTTAGCTCTGCAATACAGTTTAGCA 894
Qy 261 AsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGluAsn 280
Db 895 AACAAAGAGATTTCGTGGTGTGGTGGGATGAGCTCTGCTCACAGGTAGAGAAAT 954
Qy 281 ValAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluVal 300
Db 955 GTTGACGAGTTACAGAGCTTGAAAGTCTGGGATGGATCAAGAAACTCTGCTCTGAGGTT 1014

Qy 301 GluAlaIleLeuGluProValIleAsnLeuThrTrpProSerGlyIleHisGlnAsn 319
Db 1015 GAAGCTATTCTCGAGCCTGTAAGAATCTGACATGGCCCAAGTGAATCCATCAGAAC 1071

RESULT 4

ACN48774
ID ACN48774 standard; cDNA; 564 BP.

AC ACN48774;

XX 02-DEC-2004 (first entry)

XX Cotton primed seed EST Clone ID: LIB3825-027-Q6-K6-G2, SEQ:3555.

XX Cotton; plant; EST; expressed sequence tag; transgenic plant; seed;
XX variety DP508; library LIB3825; molecular tag; molecular marker;
XX genetic mapping; molecular mapping; seed germination; plant growth;
XX plant quality; plant yield; plant breeding; tissue printing; ss.

XX Gossypium hirsutum.

XX US2004123340-A1.

XX 24-JUN-2004.

XX 12-DEC-2001; 2001US-00021323.

XX 14-DEC-2000; 2000US-0255619P.

XX (DEIK/) DEIKMAN J.
XX (FENG/) FENG P C C.
XX (FINC/) FINCHER K L.
XX (ZIEG/) ZIEGLER T B.

XX Deikman J, Feng PCC, Fincher KL, Ziegler TE;

XX WPI, 2004-479808/45.

XX New isolated nucleic acid molecule that encodes a plant protein or its
XX fragment, useful for isolating a variety of agronomically significant
XX genes associated with plant growth, quality or yield, and as molecular
XX tags to map genes.

XX Claim 1; SEQ ID NO 3555; 34pp; English.

XX The invention relates to 17880 cotton expressed sequence tags (ESTs;
XX ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated
XX from primed or non-primed seeds from variety DP508, mature seeds from
XX variety Coker 312 Boswell 96 Field, and androecium tissue, gynoecium
XX tissue, developing fibres, carpel walls and septa from variety
XX Nuotton33B. The invention also relates to substantially purified
XX proteins or their fragments encoded by nucleic acid molecules of the
XX invention, and to transformed plants having a nucleic acid construct
XX comprising a nucleic acid of the invention. The cotton ESTs are useful as
XX molecular tags to isolate genetic regions, to isolate genes, to map
XX genes, to determine gene function and to determining whether genes are
XX members of a particular gene family. The nucleic acid molecules may be
XX used for isolating a variety of agronomically significant genes
XX associated with plant growth, quality, yield, and could also serve as
XX links in metabolic and catabolic pathways. The nucleic acid molecules are
XX also useful for identifying genes important in initiating and maintaining
XX seed germination or that may be used to mitigate stresses encountered
XX during seed germination. The ESTs additionally enable the acquisition of
XX promoters and cis-regulatory elements which will be useful to express
XX agronomically significant genes in these tissues and/or other tissues,
XX and also permits the acquisition of molecular markers useful in breeding
XX schemes, genetic and molecular mapping, and in cloning of agronomically
XX significant genes. The nucleic acid molecules are further useful for
XX detecting the expression level or pattern of a protein or mRNA and for
XX detecting the presence or quantity of a protein by tissue printing. The
XX present sequence represents a specifically claimed EST isolated from a
XX cotton variety DP508 primed seed cDNA library (LIB3825). The sequence

PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.

XX
XX
XX PA (PEKE) PE CORP NY.
XX
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX

PT genes from *Drosophila* and for elucidating cell signaling and cell-cell

XX
PS
XX

the invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from *Drosophila*. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (AB857737-
CC ABB72072). The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1038 BP; 263 A; 257 C; 304 G; 214 T; 0 U; 0 Other:

Alignment Scores:	
Pred. No.:	2.03e-53
Score:	593.50
	Length: 1038
	Matches: 128

```

DB:
Query Match: 4
Best Local Similarity: 36.66%
Query Match: 36.66%
Best Local Similarity: 42.52%
Reference Similarity: 42.52%
MisMatches: 99
Indels: 11
Gaps: 7

```

US-10-606-300-11 (1-319) x ABL20521 (1-1038)

Qy 3 LysIleGluLeuArgAlaLeuGlyValAsnThrGlyLeuLysValSerAlaValGlyPheGly 22

Accession	Protein	Accession	Protein	Accession	Protein
Db	CGEATGGAATATCGCAATCTCGGAAAGACCGGGCTGCAAGTCTCGAAAGTCTCTTTTCGGA	Qy	AlaSerProLeuGlySerValPheGlyProValAlaGluAspAspAlaValAlaThrVal		

Db	121	GGCGGGCCCTGTGCGCGAACTACGGTTTTGATTTGGAG---	GAGGGTATTAAAACTGTG	177
Qy	43	ArgGluAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyrTyrGlyGlyThr		62

Db	178	63	Qy
	CACGAGGCGGTAAGTCAAGTCAGGCATCAACTACATTGACACCTGCTCCCTGGTATGGTCAGGGT	LeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAspTyr	237

Db	238	CGCTCTGAGGAGGTCCTGGGACTGGCCCCATAAG-----GATGTGCCGGCGGAATCCTAC	291
Qy	83	IleValAlaThrIysCysGlyArgTyrLys-----GluclyPheAspPheSer	98

Db	292	TATATCGCCACGAAAGTGCCTCGTACGAACTGGACTACGATAAAATGTTTGACTTTAGT	351
Qy	99	AlaGluArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrVal	118

Db	352	GCC	AAG	AAG	CGC	GGA	AGC	TTG	AA	CT	TCT	GGC	CTG	ACT	TAC	GT	411					
Qy	119	Asp	Ile	Leu	His	Cys	His	Asp	Ile	Glu	Phe	Gly	---	Ser	Leu	Asp	Gln	Ile	Val	Ser	Glu	137

Db	412	GAGTGCATCCAGATTACGATATCGAGTTCGCCAAGGATCTGGACATTGTGATCAACGAG	471
Qy	138	ThrIleProAlaLeuGlnLysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThr	157

Db 472 A C A C T G C C C A C T T T G G A G C A G C T G G T C A A G G A G G C A A G G C A A G A T T C A T T G G A G T C C 531

CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from the US patent office at
CC seqdata.uspto.gov/sequence.html?DocID=US20040123340
XX
SQ Sequence 564 BP; 156 A; 97 C; 135 G; 175 T; 0 U; 1 Other;

Alignment Scores:	
Pred. No.:	1.05e-73
Score:	779.50
Percent Similarity:	88.83%
Best Local Similarity:	82.45%
Query Match:	88.15%
DB:	13
Gaps:	1
Indels:	2
Mismatches:	20
Conservative:	12
Matches:	155
Length:	564

US-10-606-300-11 (1-319) x ACN48774 (1-564)

Qy	66	LysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAspTyrIleValAla	85
		::: :::	
Dd	3	CGANAGCTTGGTAAAGGACTTAAAGCTCTTGGAGTTCCTAGATGAATPATAGTTTC	59

Qy 86 ThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGluArgValArgLysSer 10
Db 60 GACAAATGTGGGAGATATCGTGAAGTTTTCATGCTGAGAGACTAATCTAAAGC 11

[illegible]

Qy 126 IleGluPheGlySerIleuAspGlnIleValSerGluThrIleProAlaIleuGlnLysLeu 14
 Dd 180 ATTGAATTCGGCTCTCTTGATCAGGTTGTGAATGAGACGATTTCGGCACTTCAAAAACTG 23

Qy 146 LysGlnGlnGlyLysThrArgPheIleGlyIleThrGlyLeuProLeuAspIlePheThr 16
 Dd 240 AAGGAAGCAGGGGAAGATTGTTTCATTGGTATCACCGGGTTGCCTTGAAATTTTACT 29

Qy 166 TyrValLeuAspArgValProGlyThrValAspValIleuSerTyrCysHisTyr 18
Db 300 TATGTCTTGATAGGTTCCACCAGGACACTGTGTGATGAATATATCATATATGCCATTAT 35

Qy 186 GlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLysSerLysGlyValGly 20
:::|||||
Db 360 AGCATTAATGATTCAACATTGCGAGGATTTATTGCCCTTACTTGAACCAAGT-GTTGGC 41

Qy	206	VallSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGlnGlyProProGluTrp	22
Db	419	GTAATCAGTGCATCTCCACTTGCTATGGGACTTCCTACTGATTGTGCCACCGAGTGG	47

Qy	HisProAlaSerProGluLeuLysSerAlaSerLysAlaValAlaHisCysLysSer
Db	CATCGGGCATCTCCGAACCTCAAGTCGCCTGCCAAGTGTCTATTATTAAAGAG

Qy 246 LysGlyLysLysIleThrLysLeu 253
 ||||| |||||
Dd 539 AAAGAAAGAATATTTCGAAGTTA 562
 ||:::||

RESULT 5
ABL20521
ID ABL20521 standard; DNA; 1038 BP.

AC	ABL20521;
XX	
DT	26-MAR-2002 (first entry)

DE *Drosophila melanogaster* genomic polynucleotide SEQ ID NO 13036.
XX
KW *Drosophila*; developmental biology; cell signalling; insecticide;

XX
OS *Drosophila melanogaster*.
XX

Qy 158 GlyLeuProLeuAspPheThrTyrValLeuAspArgValProProGlyThrValAsp 177
 Db 532 GCTTACCGATTTCGGTCTTAAGGAGTTCCTGACCCGAACA--GCCGGAAGACTCGAT 588
 Qy 178 VallLeuSerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuPro 197
 Db 589 ACGTCTCCTACCTATGCCAGATACACCTGACCGATGAACGCTCTCGAGTACCTGGAT 648
 Qy 198 TyrLeuLysSerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeu 217
 Db 649 TTCCTCAAGTCCAGAACTGGGCTCATCTGTGCGCAGCTCATGCTCGACTGCTG 708
 Qy 218 ThrGluGlnGlyProProGluTrpHisProLaserProGluLeuLysSerAlaSerLys 237
 Db 709 ACCAATCGCGTCCACAGCCATGGCATCCGGCCAGTGATGACGAGAGGCCATGTCGCGG 768
 Qy 238 AlaAlaValAlaHisCysLysSerLysGlyLysIleThrLysLeuAlaLeuGlnTyr 257
 Db 769 AAGCATCGGAGTCTGCAAGGAACGCGCGTGGAGCTGGGCAAGCTGGCCATGTAATAC 828
 Qy 258 SerLeuAlaAsn--LysGluIleSerSerValLeuValGlyMetSerSerValSerGln 276
 Db 829 ACGATGAGCGGACTCCCGAAGTGAACCTTCCTTAACGGGCGATGCACAGCGCCAGTTG 888
 Qy 277 ValGluGluAsnValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThr 296
 Db 889 CTGCGAATCACTGGATGCC--AACGAAGTGGGCTCAGCGATAGAGGAGGAGGAGTG 945
 Qy 297 Leu 297
 Db 946 TTG 948

RESULT 6

ABL20520/c
 ID ABL20520 standard; DNA; 3724 BP.

XX AC ABL20520;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 13033.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical; gene; ds.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX PR 23-MAR-2000; 2000US-0191637P.

XX PR 11-JUL-2000; 2000US-00614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.

XX PS Claim 1; SEQ ID NO 13033; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA sequences (AB101840-ABL16175) and the encoded proteins (ABB57737-CC ABB72072). The sequence data for this patent did not form part of the CC printed specification, but was obtained in electronic format directly CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 3724 BP; 985 A; 815 C; 802 G; 1122 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.39e-47 Length: 3724
 Score: 546.50 Matches: 131
 Percent Similarity: 51.19% Conservative: 63
 Best Local Similarity: 34.56% Mismatches: 97
 Query Match: 33.76% Indels: 88
 DB: 4 Gaps: 9

US-10-606-300-11 (1-319) x ABL20520 (1-3724)

Qy 3 LysIleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGlyPheGly 22
 Db 2664 CGGATGGAATATCGAATCTCGAAAGACCGCCTGCAAGTCTCGAAAGTCTCTTTCCGA 2605
 Qy 23 -AlaSerProLeuGlySerValPheGly- 31
 Db 2604 GCGCGCGCCCTGTGCGCAACTACGGGTAAAGAGAAATCTTAACAATCAATTACTCCCG 2545
 Qy 31 - 31
 Db 2544 TACACAGAAAGAAATATTTCTTTCTTGGCATTAAATAATAAATTTCTAGAAAGATTG 2485
 Qy 31 - 31
 Db 2484 CTGAAAGTACGAATTCGTAAATTAATAATGCGGAAATGTTTGAGATTGTATTCACT 2425
 Qy 32 - - - - -ProValAlaGluAsp- - - - -AspAlaValAlaThrValArgG 44
 Db 2424 CTAATTTTCATATATTCCTCTTCAGTTTGTGAGAGGAGGTATTAACACTGTGCACGA 2365
 Qy 44 uAlaPheArgLeuGlyIleAsnPheAspThrSerProTyrTyrGlyGlyThrLeuSe 64
 Db 2364 GCGCGTAAAGTCAGGCATCAACTACATTCAGTCTCCCTGGTATGTCAGGTCGCTC 2305
 Qy 64 rGluLysMetLeuGlyLysGlyLeuLeuAlaLeuGlnValProArgSerAspTyrIleVa 84
 Db 2304 TGAGGAGTCTCTGGACTGGCCCTAAAG- - - - -GATGTGCGCGGGAATCTCTACTAT 2251
 Qy 84 lAlaThrLysCysGlyArgTyrLys- - - - -GluGlyPheAspPheSerAlaG 100
 Db 2250 CGCCACGAAAGTGGCTCGCTACGAACCTGGACTACGATAAAATGTTTGTAGTGCCAA 2191
 Qy 100 uArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAsp 120
 Db 2190 GAAGACGCGCGAAAGCGTGGAGAGAGCTTGAACACTCTTGGCTGGAGTACGTTGATG 2131
 Qy 120 eLeuHisCysHisAspIleGluPheGly- - -SerLeuAspGlnIleValSerGluThr 139
 Db 2130 CATCCAGATTTCAGATATCGAGTTCGCCAAGGATCTGGACATTGTGTCAACAGACACT 2071
 Qy 139 eProAlaLeuGlnLysLeuLysGlnGlyLysThrArgPheIleGlyIleThrGlyLe 159
 Db 2070 GCGCCACTTTGGAGCAGCTGGTCAAGGAGGCGCAAGAGATTCATTGGAGTGTCCGCTTA 2011
 Qy 159 uProLeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAspVal- 178
 Db 2010 CCCGATTTCCGTGTCTTAAGGAGTTCCTGACCCGAAACA- - -GCCGGAAGACTCGATGTG 1954
 Qy 179 - - - - -IleLeu 180
 Db 1953 TTTATAGCATACACTTCAAAATTTCAAGTTTAAATCTCATATCGCCACCAGCGTCTC 1894
 Qy 180 uSerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeu 200

Db 1893 CACCTATGCCAGATACACCCCTGACCGATGAACCGCTCTCGGAGTACCTGGATTTCTTCAA 1834
QY 200 sSerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluG1 220
Db 1833 GTCCAGAACCTGGGGCGTTCATCTGTGCGCGAGCTCATGCTCGGAGCTGCTGACCAATGC 1774
QY 220 nGlyProGluTTPHisProAlaSerProGluLeuLysSerAlaSerLysAlaAlaVa 240
Db 1773 CGGTCCACAGCCATGGCATCCGCGCAGTGTATGAGCAGAGGCCATTGCCCGAAGGCATC 1714
QY 240 lAlaHisCysLysSerLysGlyLysIleThrLysLeuAlaLeuGlnTyrSerLeuAl 260
Db 1713 GGAGTCTGCAAGGAACGGCGGTGGAGCTGGGCAAGCTGGCCATGCTACTACACGATGAG 1654
QY 260 aAsn---LysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluG1 279
Db 1653 CGGACTGCCCGAAGTGGACACCTCTCAACGGGCGATGCAGACGCGCCAGTTGCTCGGAAT 1594
QY 279 uAsnValAlaAlaValThrGluLeuGluSerLeuGlyMetaspGlnGluThrLeu 297
Db 1593 CAACTGGATGCC---AACGAAGTGGGCTCAGCGATTAAGGAGCAGGAAGTGTG 1542

RESULT 7

ABL28825
ID ABL28825 standard; DNA; 1051 BP.
XX ABL28825;
AC ABL28825;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 37948.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
FN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
DR
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 37948; 21pp + Sequence Listing; English.
XX

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 1051 BP; 242 A; 271 C; 305 G; 233 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3.62e-48 Length: 1051

RESULT 8

Score: 544.50 Matches: 117
Percent Similarity: 58.75% Conservative: 71
Best Local Similarity: 36.56% Mismatches: 117
Query Match: 33.63% Indels: 15
DB: 4 Gaps: 7
US-10-606-300-11 (1-319) x ABL28825 (1-1051)
QY 3 LysIleGluLeuArgAlaLeuGlyAsnThrGlyLysValSerAlaValGlyPheGly 22
Db 83 CGTATGGAGTACCGCAACTCGGATCCACGGGCTCTCGAGTATCCAAATATGCTTTGGGT 142
QY 23 AlaSerProLeuGlySerValPheGly---ProValAlaGluAspAspAlaValAlaThr 41
Db 143 GGTGCCACTCTCTCAAGCTCTTCTCCGATGACTTCGAGTCGGAGAGGGCATCTCTCAG 202
QY 42 ValArgGluAlaPheArgLeuGlyIleAsnPheAspThrSerProTyrTyrGlyGly 61
Db 203 GTGCAGGAAGCCATTAGATCCGGTATCAACTACATAGACACGGCTCCCTTTTATGGCAA 262
QY 62 ThrLeuSerGluLysMetLeuGlyLysGlyLysAlaLeuGlnValProArgSerAsp 81
Db 263 GGCAAATCGGAAGAGCTGTTGGCCAGCGCTCAAG-----GATGTGCCCGGAGGCC 316
QY 82 TyrIleValAlaThrLysCysGlyArgTyrLys-----GluGlyPheAspPhe 97
Db 317 TATTATATAGCAACTAAAGTTGCAGTTACGAGTTGGATGCCAAACAATATGTTTCGACTAT 376
QY 98 SerAlaGluArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyr 117
Db 377 ACGGTGCCAAGGCTCGGAGAGTGTGAAGCGTAGTCTGGAGCTGCTCCAGTTGCACAGG 436
QY 118 ValAspIleLeuHisCysHisAspIleGluPheGly---SerLeuAspGlnIleValSer 136
Db 437 GTGGACGTACTACAGGTTTCATGAGTTCGAGTCCGACCTAGTCTGGACATGCTGTGTAAT 496
QY 137 GluThrIleProAlaLeuGlnLysLeuLysGlnGlyLysThrArgPheIleGlyIle 156
Db 497 GAGACCATACCCGCTCTCGAGGAGTACGTCAGGGGGGAAAGGCTCGATTATTCGAGTTC 556
QY 157 ThrGlyLeuProLeuAspIlePheThrTyrValLeuAspArgValProGlyThrVal 176
Db 557 ACCGCTACGATGTGACGTGCTGAAGGAGTGTGCGAGCGG---GGCAAGGTCCGATC 613
QY 177 AspValIleLeuSerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeu 196
Db 614 CAGGTGTGTCTCAACTATGCTCCGCTACACCTGTTTGGACAACACTTTCTGCGCCACATG 673
QY 197 ProTyrLeuLysSerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeu 216
Db 674 AAGGCTTCCAGGAGATGGCGTGTGCGCTTGTGTCGGCGCGGCATCTCTCTGGAGTCTC 733
QY 217 LeuThrGluGlnGlyProGluTrpHisProAlaSerProGluLeuLysSerAlaSer 236
Db 734 CTAAGCAACGCTGGACCCCGCTGTCGTCATCTCTGTCGTCGTCGTCGTCGTCGTCGTC 793
QY 237 LysAlaAlaValAlaHisCysLysSerLysGlyLysIleThrLysLeuAlaLeuGln 256
Db 794 AAAAGGGGAGCGCAAAATCTGTCAAGAGGAGCAACGTTGAGCTTGGAAAGCTGGCCATG 853
QY 257 TyrSerLeuAlaAsnLysGluIleSerValLeuValGlyMetSerSerValSerGln 276
Db 854 TATACATATGCACTGGATGGGCGGCCACCTTCTCATCGTATCCCAACCGCAAGCTG 913
QY 277 ValGluGluAsnValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThr 296
Db 914 CTGCGGATTAACCTTGGACGCGATCTTCAC-----GGTCTCACTCTCCACGCA 961
QY 297 LeuSerGluValGluAlaIleLeuGluProVal-----LysAsnLeuThrTrpProSer 314
Db 962 CAGGAAGTGTGCAGTATCTGCGGAAACGTTTTTACCAAGTCTCTACAGTTGGGGTCC 1021

ABL28824/c
 ID ABL28824 standard; DNA; 3308 BP.
 XX AC ABL28824;
 XX DT 26-MAR-2002 (first entry)
 XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 37945.
 XX KW Drosophila; developmental biology; cell signalling; insecticide;
 XX KW pharmaceutical; gene; ds.
 XX OS Drosophila melanogaster.
 XX PN WO200171042-A2.
 XX PD 27-SEP-2001.
 XX PP 23-MAR-2001; 2001WO-US009231.
 XX PR 23-MAR-2000; 2000US-0191637P.
 XX PR 11-JUL-2000; 2000US-00614150.
 XX XX (PEKE) PE CORP NY.
 XX PA
 XX PI Venter JC, Adams M, Li PWD, Myers EW;
 XX XX
 XX DR WPI; 2001-656860/75.
 XX
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 XX PT genes from Drosophila and for elucidating cell signaling and cell-cell
 XX PT interactions.
 XX PS Claim 1; SEQ ID NO 37945; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
 CC ABBS70272). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX SQ Sequence 3308 BP; 898 A; 727 C; 732 G; 951 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1,68e-44 Length: 3308
 Score: 517.00 Matches: 118
 Percent Similarity: 55.49% Conservative: 69
 Best Local Similarity: 35.01% Mismatches: 115
 Query Match: 31.93% Indels: 36
 DB: 4 Gaps: 7

US-10-606-300-11 (1-319) x ABL28824 (1-3308)

QY 3 LysileGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGlyPheGly 22
 Db 2226 CGTATGGAGTACCGCAACTCGATCCAGGGTCTCGAGTATCCAAATTTGTTGGT 2167
 QY 23 AlaSerProLeuGlySerValPheGly---ProValAlaGluAspAlaValAlaThr 41
 Db 2166 GTGGCCACTCTCTCCAAAGCTCTCTCCGATGATCTTCGATGAGTCCGAGGAGGCACTCTCAG 2107
 QY 42 ValArgGluAlaPheArgLeuGlyIleAsnThrPheAspThrSerProTyrTyrGlyGly 61
 Db 2106 GTGCAGGAGCCATTAGATCCGGTATCACTACATAGACACGGCTCCCTTTATGCGCAA 2047
 QY 62 ThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAsp 81
 Db 2046 GCGAAATCGGAAGAGTGTGTCGCCAGCGGCTCAAG-----GATGTGCCCGGAGGCC 1993

QY 82 TyrIleValAlaThrLysCysGlyArgTyrLys-----GluGlyPheAspPhe 97
 Db 1992 TATTATATAGCAACTAAAGTTGCACGTTACGAGTTGGATCCAAACAATATGTCGACTAT 1933
 QY 98 SerAlaGluArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyr 117
 Db 1932 ACGCTCCCAAGGCTCGGAGAGTGTGAAGCGTAGTCTGGAGCTGCTCCAGTTGGACAGG 1873
 QY 118 ValAspIleLeu-HisCys----- 123
 Db 1872 GTGGACGTACTACAGGTGTGTTGGTATACCAATATAGATTTGCTATATAACTAACAC 1813
 QY 124 -----HisAspIleGluPheGly---SerLeuAspGlnIleValSerGluTh 138
 Db 1812 CTTGTGACAGGTTTCATGACGTGGATGGCGCACCTAGTCTGACATGTTGCTGAATGAGAC 1753
 QY 138 rIleProAlaLeuGlnLysLeuLysGlnGluGlyThrArgPheIleGlyIleThrG1 158
 Db 1752 CATACCCGTCCTCGAGGAGTACGTCAGGCGGGAAGGCTCGATTCATCGAGTCCCGC 1693
 QY 158 yLeuProLeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAspVa 178
 Db 1692 CTACGATGTGACGTGCTGAAGGAGTGTCCGAGCGG---GCAAAAGTCCGATCCAGT 1636
 QY 178 lIleLeuSerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTy 198
 Db 1635 GGTGCTCAACTATGCCCGCTACACCTGTTGGACAACACTTTGCTGCGCCACATGAAGGC 1576
 QY 198 rLeuLysSerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuTh 218
 Db 1575 CTTCCAGGAGATGGCGGTGCTGTGTGCGCGCGCGCACTCTCTGGGACTCCTTAAG 1516
 QY 218 rGluGlnGlyProGluTyrHisProAlaSerProGluLeuLysSerAlaSerLysAl 238
 Db 1515 CRAAGCTGGACCCAGCTCTGGATCTCTGGTAGTCCGGAATCTTAGTGTGGGCAACG 1456
 QY 238 aAlaValAlaHisCysLysSerLysGlyLysIleThrLysLeuAlaLeuGlnTyrSe 258
 Db 1455 GGGAGCCGMAATCTGTCAAGAGAGAGACGTTGAGCTTGGAAAGCTGGCCATGACTATAC 1396
 QY 258 rLeuAlaAsnLysGluLysSerValLeuValGlyMetSerSerValSerGlnValG1 278
 Db 1395 AATCAACTGGATGGCGCGCCACTCTCTCATCGGTATCCCAACCCGAAAGCTGCTGG 1336
 QY 278 uGluAsnValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSe 298
 Db 1335 GATTAACCTGGACCGGATC-----TTGACGGTCTCACTTCCCA 1297
 QY 298 rGluValGluAlaIleLeuGluProValLysAsnLeuThrTrpProSer 314
 Db 1296 CGAACAGGAAGTGTGTGCA-GTATCTGCGCGGAAAGTAAGTTGGCCAACA 1249
 RESULT 9
 ACN48848/c
 ID ACN48848 standard; cDNA; 563 BP.
 XX AC ACN48848;
 XX DT 02-DEC-2004 (first entry)
 XX DE Cotton primed seed EST Clone ID: L1B3825-027-Q6-N6-G2, SEQ:3629.
 XX KW Cotton; plant; EST; expressed sequence tag; transgenic plant; seed;
 XX KW variety D50B; library L1B3825; molecular tag; molecular marker;
 XX KW genetic mapping; molecular mapping; seed germination; plant growth;
 XX KW plant quality; plant yield; plant breeding; tissue printing; ss.
 XX OS Gossypium hirsutum.
 XX PN US2004123340-A1.
 XX PD 24-JUN-2004.

[illegible]

Db	441	GCCTGCCAAGCTGCTGCTATATTTAAAGAGAAAGAAATATTTTCGAAGTTAGCT	382
Qy	255	LeuGlnTyrSerLeuAlaAAsnLYsGluIleSerSerValLeuValGlyMetSerSerVal	274
Db	381	ATGCAATACAGCTTGTCACAAAGATATTTTCGACAGTGTGGTGGCATGAACTCGGTT	322
Qy	275	SerGlnValGluGluAenValAlaAaValThrGluLeuGluSerLeuGlyMetAspGln	294
Db	321	AAACAGGTTGAAGAGAGATGTTGCCGCTGCAACAGAACTTGCACTATTTCGGGAAAGATCAT	262
Qy	295	GluThrLeuSerGluValGluAlaIleLeuCluP-roValLYsAsnLeuThrTrpProSer	314
Db	261	GAAACTCTAGCTGAGGTGGAGCAATACTGAAGCCAGTGAAGAAATCAGACATGGCCGAGC	202
Qy	315	GlyIleHisGlnAen 319	
Db	201	GGATTTCAACGAGC 187	
RESULT 10			
ABLO6602/c			
ID	ABLO6602	standard; cDNA; 3014 BP.	
AC	ABLO6602;		
XX			
DT	26-MAR-2002	(first entry)	
XX			
DE	Drosophila melanogaster	expressed polynucleotide SEQ ID NO 14288.	
XX			
KW	Drosophila; developmental biology; cell signalling; insecticide;		
XX	pharmaceutical; gene; ss.		
XX			
OS	Drosophila melanogaster.		
XX			
PN	WO200171042-A2.		
XX			
PD	27-SEP-2001.		
XX			
PF	23-MAR-2001; 2001WO-US009231.		
XX			
PR	23-MAR-2000; 2000US-0191637P.		
PR	11-JUL-2000; 2000US-00614150.		
XX			
PA	(PEKE) PE CORP NY.		
XX			
PI	Venter JC, Adams M, Li PWD, Myers EW;		
XX			
DR	WPI; 2001-656860/75.		
DR	P-PSDB; ABB62499.		
XX			
PT	New isolated nucleic acid detection reagent for detecting 1000 or more		
PT	genes from Drosophila and for elucidating cell signalling and cell-cell		
PT	interactions.		
XX			
PS	Claim 1; SEQ ID NO 14288; 21pp + Sequence Listing; English.		
XX			
CC	The invention relates to an isolated nucleic acid detection reagent		
CC	capable of detecting 1000 or more genes from Drosophila. The invention is		
CC	useful in developmental biology and in elucidating cell signalling and		
CC	cell-cell interactions in higher eukaryotes for the development of		
CC	insecticides, therapeutics and pharmaceutical drugs. The invention		
CC	discloses genomic DNA sequences (ABU16176-ABL30511), expressed DNA		
CC	sequences (ABLO1840-ABU16175) and the encoded proteins (ABB5773-		
CC	ABB72072). The sequence data for this patent did not form part of the		
CC	printed specification, but was obtained in electronic format directly		
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences		
XX			
SQ	Sequence 3014 BP; 852 A; 644 C; 610 G; 908 T; 0 U; 0 Other;		
Alignment Scores:			
Pred. No.:	3e-38	Length:	3014
Score:	458.00	Matches:	116
Percent Similarity:	54.30%	Conservative:	67

QY 63 LeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAspTyr 82
DB 175 CTT : : : : :
QY 83 IleValAlaThrLysCysGlyArgTyr : : : : :
DB 211 TATATTGCATTAAGTGGCGGCTACGGGTGGATCCGAAGATATATTGACTATTCG 270
QY 99 AlaGluArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrVal 118
DB 271 GCTGACAAAGCTCGGAGAGTGTGAAGCGGAGTCTGGAGCGGCTCCAGTTGGACAGGGTG 330
QY 119 AsplleLeuHisCysHisAsplleGluPheGly : : : : :
DB 331 GACATACTACAGTTTCATGACGTGACGCGGCGGACCAATCTAGACATATGCTGAATGAG 390
QY 138 ThrIleProAlaLeuGlnLysLeuLysGlnGluLysThrArgPheIleGlyIleThr 157
DB 391 ACCATACCGTCTCGAGAGTACGTCCAGGGGGAAGGCTCGATTTCATCGAGTCACC 450
QY 158 GlyLeuProLeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAsp 177
DB 451 GCCTACGATGTGAGCGTGTGAAGAGTGTGCGGAGCGG---GGCAAGGCTCGCATTCAG 507
QY 178 VallileLeuSerTyrCysHisTyrGlyValIleAspSerThrLeuLeuAspLeuPro 197
DB 508 GTGTGTCTCAACTATGTCTGTGTACACCTTTTGTAGACAAACACCTTGTCTGCTACATGAAG 567
QY 198 TyrLeuLysSerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeu 217
DB 568 GACTTCCAGAAATGGAGTGGGCGTGTGTCTGCGGCGCTCACTCATTTGGACTCTTG 627
QY 218 ThrGluGlnGlyProGluTrpHisProAlaSerProGluLeuLysSerAlaSerLys 237
DB 628 AGAAACGCTGGACCATGTCATCGCATCCCGGTAGTTCAGGAAATCCTGCGCGTGGCCAAA 687
QY 238 AlaAlaValAlaHisCysLysSerLysGlyLysIleThrLysLeuAlaLeuGlnTyr 257
DB 688 CGGGGGCGGAAATCTGCGACGAGGAAACGTGGAGCTGGGAAAGCTGGCATGTACTAT 747
QY 258 SerLeuAlaAsnLysGluIleSerValLeuValGlyMetSerValSerGlnVal 277
DB 748 ACGATGCACTGGATGGGCGGCGCACTTCCTTCATCGGCATCCCAACGGAAGCTGTG 807
QY 278 GluGluAsnValAlaValThrGluLeuGluSerLeuGlyMet : : : : :
DB 808 CGGATTAACCTGGACGCGATCTTCGAC : : : : :
QY 295 GluThrLeu 297
DB 856 GAAGTGTGCTG 864

RESULT 12

AAQ62089

ID AAQ62089 standard; DNA; 987 BP.

XX AC AAQ62089;

XX DT 16-OCT-2003 (revised)

XX DT 14-DEC-1994 (first entry)

XX DE L-fucose dehydrogenase gene.

XX KW L-fucose dehydrogenase; L-FDH; enzyme; ss.

XX OS Pseudomonas sp; No.1143.

XX PN JP06090765-A.

XX PD 05-APR-1994.

XX PF 11-SEP-1992; 92JP-00243372.

XX

PR 11-SEP-1992; 92JP-00243372.
XX (KIKK) KIKKOMAN CORP.
PA (NODA) ZH NODA SANGYO KAGAKU KENKYUSHO.
XX WPI; 1994-146990/18.
DR P-FSDB; AAR51284.
XX Novel L-fucose dehydrogenase gene - a new recombinant DNA and the prepn.
PT of L-FDH using a L-FDH vector.
XX Claim 1; Page 4-5; 6pp; Japanese.
XX The L-FDH gene is new and can be used for the recombinant prodn. of the
CC enzyme. (Updated on 16-OCT-2003 to standardise OS field)
XX SQ Sequence 987 BP; 140 A; 371 C; 340 G; 136 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 4.13e-29 Length: 987
Score: 366.00 Matches: 103
Percent Similarity: 49.25% Conservative: 62
Best Local Similarity: 30.75% Mismatches: 122
Query Match: 22.61% Indels: 48
DB: 2 Gaps: 9
US-10-606-300-11 (1-319) x AAQ62089 (1-987)
QY 8 AlaLeuGlyAsnThrGlyLeuLysValSerAlaValGlyPheGlyAlaSerProLeuGly 27
DB 19 GCGGGCGGCTGCGGCGCTGCCCATCCGCGCCCTCGGCTACGGCGCGCCCAACGTGGGC 78
QY 28 SerValPheGlyProValAlaGluAspAlaValAlaThrValArgGluAlaPheArg 47
DB 79 AACCTCTTCCGCGGCTGAGCGATGACGAGGCTGGGCGTCTCGAGCGCGCTGGGAT 138
QY 48 LeuGlyIleAsnPheAspThrSerProTyrTyrGlyThrLeuSerGluLysMet 67
DB 139 GCGGCGATCCCTATTACGACCGCGCGCTACGCGGCTCGGGCTGAGCGAGAGCGC 198
QY 68 LeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAspTyrIleValAlaThrLys 87
DB 199 CTGGTGCCTTCTTCGAGACC-----AAGCGCGCGAGAGTTCGTCTGTGCGACCAAG 252
QY 88 CysGlyArg----- : : : : :
DB 253 GCGGGCGCTGTCTGCGCCGAGACCCCGAGCGCGCGGCTCGACACCGACAAC 312
QY 91 -----TyrLysGluGlyPheAspPheSerAlaGluArgValArg 103
DB 313 GACTTCCACGTGCGCGACGACTGCGCGCGAGTGGGACTTCACGAGCAGGGCATCCGT 372
QY 104 LysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIleLeuHisCys 123
DB 373 GCGAGCATGCGGAGTCTGAGAGCGGCTCGGCGCTCGACCGCATCGACTGCTGTACCTG 432
QY 124 HisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIleProAlaLeuGln 143
DB 433 CACGACCCGAGCGGCGACGACTCGACCTCGCGCTCGCGCTTCCCGCGCTCGAG 492
QY 144 LysLeuLysGlnGluGlyThrArgPheIleGlyIleThrGlyLeuProLeuAspIle 163
DB 493 AAGGTGCGCGCGGCGGCGTGTGAGCGCATCGGCTCGATCGATGCTCGGATGCC 552
QY 164 PheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeuSerTyrCys 183
DB 553 CTGACCCGGCG-----GTCCGCGAGCGGAGCTCGACCTCATCATGTGCTCGCGG 603
QY 184 HisTyrGlyValAsnAspSerThrLeuLeu-----AspLeuLeuPro 197
DB 604 CGCTAC-----ACGTGCTCGAGCAGCGCGCGCGCGCGAGGTGTGCT 648
QY 198 TyrLeuLysSerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeu 217

Db 649 GCATGCGCTGAGAACGCCACCGGATCGCGGCATCCGCTTCAATTCGGGACTGCTC 708
Qy 218 ThrGluGlnGlyProPro-----GluTrpHisProAlaSerProGluLeu 232
Db 709 GCGCAGCGGAGCGGAGCGGCGGCGCTACGAGTACGCTCGCTCGCTCGCTCGCTG 768
Qy 233 LysSerAlaSerLysAlaAlaValAlaHisCysLysSerLysGlyLysLysLysLys 252
Db 769 TGGGATCGCTGCTGGTGGATCGCGGATCGCGCAACCGACGAGTACCGCTCGCGGCG 828
Qy 253 LeuAlaLeuGlnTyrSerLeuAlaAsnLysGluLysSerValLeuValGlyMetSer 272
Db 829 GCGCGGATCCAGTTCCTCCCTCGAGTCCGCTCGCTCGCTCGCTCGCTCGCTCG 888
Qy 273 SerValSerGlnValGluGluAsnValAla---AlaValThrGluLeuGluSerLeuGly 291
Db 889 CGCCCGCGGAGCTGACCGAGACGCCGAGTACCGCGGCTCGAGATCCCGGCC---GGG 945
Qy 292 MetAspGlnGluThrLeuSerGluValGluAlaLeuLeuGluPro 306
Db 946 CTGTGGCGCGAG---CTGGCGGAGCGCGCTGATCCCGCCCGCC 987

RESULT 13

ABL20064/c
ID ABL20064 standard; DNA; 10400 BP.

AC ABL20064;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 11665.

XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.

XX Claim 1; SEQ ID NO 11665; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
XX ABBS70272). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 10400 BP; 2886 A; 2139 C; 2340 G; 3035 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 4,68e-25 Length: 10400
Score: 342.00 Matches: 76
Percent Similarity: 58.03% Conservative: 36
Best Local Similarity: 39.38% Mismatches: 54
Query Match: 21.12% Indels: 27
DB: 4 Gaps: 5

US-10-606-300-11 (1-319) x ABL20064 (1-10400)

Qy 3 LysileGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGlyPheGly 22
Db 592 CGTATGAGTACCGCCCACTCGGATCCACGGTCTGGAGTATCCAAATGCTTTGGST 533
Qy 23 AlaSerProLeuGlySerValPheGly---ProValAlaGluAspAlaValAlaThr 41
Db 532 GGTGCCACTCTCTCCAAGCTCTTCGGATCGACTTCGATCGCGAGGAGGCATCTCTCAG 473
Qy 42 ValArgGluAlaPheArgLeuGlyLysAsnPhePheAspThrSerProTyrTyrGlyGly 61
Db 472 GTGCGAGGAGCCATTAGATCCGGTATCACTACATAGACACGGCTCCCTTTATGGCCAA 413
Qy 62 ThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAsp 81
Db 412 GCGAATCGGAGAGCTGCTTGGCCAGCGCTCAAG-----GATGTCCCGGAGGCC 359
Qy 82 TyrileValAlaThrLysCysGlyArgTyrLys-----GluGlyPheAspPhe 97
Db 358 TATTATATAGCACTAAAGTTGCACGTTACGATGGATCCAAACAATATGTTGACTAT 299
Qy 98 SerAlaGluArgValArgLysSerLysSerLysSerLysGluArgLeuGluLeuAspTyr 117
Db 298 ACGCTCCCAAGCGCTCGGAGAGTGTGAAGCGTAGTCTGGAGCTGCTCCAGTTGGACAGG 239
Qy 118 ValAspIleLeu-HisCys-----GATGTCCCGGAGGCC 123
Db 238 GTGACGCTACTACAGGTGTGTTGGTATACCAATATATAGATTGCTATAAATAACAC 179
Qy 124 -----HisAspIleGluPheGly---SerLeuAspGlnIleValSerGluTh 138
Db 178 CTTGTGACAGGTTTCATGACGTGGATCGGCACCTAGTCTGGACATGGTGTGAATGAGAC 119
Qy 138 rIleProAlaLeuGlnLysLeuLysGlnGlyLysThrArgPheIleGlyIleThrGl 158
Db 118 CATACCCGTCCTCGAGGAGTACGTCAGCGGCGGAAAGGCTCGATTCATCGGAGTCAACGC 59
Qy 158 yLeuProLeuAspIlePheThrTyrValLeuAspArg 170
Db 58 CTACGATGTGACGCTGCTGAAGAGGAGTGTGCGGAGCGG 22

RESULT 14

ABL28814/c

ID ABL28814 standard; DNA; 10418 BP.

XX ABL28814;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 37915.

XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX


```
Qy 62 ThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAsp 81
Db 412 GGCAAAATCGGAAGAGCTGCTTGGCCAGCGCTCAAG-----GATGTGCCCCGGGAGGCC 359
Qy 82 TyrIleValAlaThrLysCysGlyArgTyrLys-----GluGlyPheAspPhe 97
Db 358 TATTATATAGCAACTAAAGTTGCAGTTACAGTTGATCCAAACAATATGTTGACTAT 299
Qy 98 SerAlaGluArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyr 117
Db 298 ACGGCTGCCAAGGCTCGGAGAGTGTGAAGGCTAGTCTGGAGCTGCTCCAGTTGGACAGG 239
Qy 118 ValAspIleLeu-HisCys----- 123
Db 238 GTGGACGTACTACAGGTGTGTTGGTATACCAATATATAGATTTGCTATAAACTAAACAC 179
Qy 124 -----HisAspIleGluPheGly---SerLeuAspGlnIleValSerGluTh 138
Db 178 CTTGTGACAGGTTTATGACGTGGATGGCGGCACCTAGTCTGGACATGGTGTGAATGAGAC 119
Qy 138 rIleProAlaLeuGlnLysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGl 158
Db 118 CATACCCGTCTCGAGGAGTACGTCCAGCGGGAAGGCTCGATTTCATCGGAGTCACCGC 59
Qy 158 yLeuProLeuAspIlePheThrTyrValLeuAspArg 170
Db 58 CTACGATGTGGACGTGCTGAAGAGAGTGTGCCGAGCGG 22
```

Search completed: November 13, 2005, 08:51:32
Job time : 636 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 13, 2005, 08:39:29 ; Search time 208 Seconds
(without alignments)

2509.482 Million cell updates/sec

Title: US-10-606-300-11

Perfect score: 1619

Sequence: 1 MTKIELRAGNLGKLVSAVG.....VEALPEVKNLTPSGIHQN 319

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPRO_pool_h/US10606300/runat_07112005_094939_8202/app_query.fasta.1.455
-DB=Issued Patents NA -QWTF=fastcap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10606300 @CGN 1.1.69 @runat_07112005_094939_8202 -NCPU=6 -ICPU=3
-NO_MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/1/ina/5A COMB.seq:*\n2: /cgn2_6/ptodata/1/ina/5B COMB.seq:*\n3: /cgn2_6/ptodata/1/ina/6A COMB.seq:*\n4: /cgn2_6/ptodata/1/ina/6B COMB.seq:*\n5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq:*\n6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1619	100.0	960	4	US-09-630-983A-12
2	351.5	21.7	2682	1	US-07-855-793-3
3	317	19.6	1104	4	US-09-252-991A-10047
c	317	19.6	2514	4	US-09-252-991A-9791
5	314.5	19.4	1044	4	US-09-902-540-3178
c	314.5	19.4	14382	4	US-09-902-540-1145
7	302.5	18.7	1026	4	US-09-252-991A-13505
c	302.5	18.7	1515	4	US-09-252-991A-13948
9	300.5	18.6	999	4	US-09-902-540-3489
c	300.5	18.6	17188	4	US-09-902-540-1166
11	296.5	18.3	999	4	US-09-107-532A-3237
12	296	18.3	1005	4	US-09-328-352-1953

13	291.5	18.0	945	3	US-09-134-001C-289	Sequence 289, App
c	291.5	18.0	2993	4	US-09-710-279-3821	Sequence 3821, Ap
15	291	18.0	1044	4	US-09-673-198-9	Sequence 9, Appli
16	291	18.0	4390	4	US-09-673-198-11	Sequence 11, Appl
17	277.5	17.1	1095	4	US-09-489-039A-5370	Sequence 5370, Ap
18	277	17.1	1002	4	US-09-724-823-24	Sequence 24, Appl
19	274	16.9	1053	4	US-09-489-039A-2381	Sequence 2381, Ap
20	271	16.7	987	4	US-09-489-039A-5632	Sequence 5632, Ap
21	270	16.7	5504	4	US-09-902-540-716	Sequence 716, App
22	264	16.3	1077	4	US-09-252-991A-14964	Sequence 14964, A
c	264	16.3	1155	4	US-09-252-991A-14787	Sequence 14787, A
24	261.5	16.2	5541	1	US-08-920-812-20	Sequence 20, Appl
c	261.5	16.2	5541	1	US-08-920-812-20	Sequence 20, Appl
26	261.5	16.2	5541	1	US-08-921-177-20	Sequence 20, Appl
c	261.5	16.2	5541	1	US-08-921-177-20	Sequence 20, Appl
28	261.5	16.2	5541	2	US-08-920-828-20	Sequence 20, Appl
c	261.5	16.2	5541	2	US-08-920-828-20	Sequence 20, Appl
30	254	15.7	1044	4	US-09-489-039A-2318	Sequence 2318, Ap
31	253	15.6	3019	4	US-09-949-016-528	Sequence 528, App
32	252.5	15.6	3911	4	US-09-586-921-61	Sequence 61, Appl
c	252.5	15.6	3911	4	US-09-586-921-61	Sequence 61, Appl
34	252.5	15.6	34953	4	US-09-902-540-5431	Sequence 5431, Ap
c	252.5	15.6	34953	4	US-09-902-540-5431	Sequence 5431, Ap
36	250	15.4	978	4	US-09-902-540-7426	Sequence 7426, Ap
c	250	15.4	978	4	US-09-902-540-7426	Sequence 7426, Ap
38	249	15.4	976	4	US-09-949-016-5100	Sequence 5100, Ap
c	249	15.4	976	4	US-09-949-016-5100	Sequence 5100, Ap
40	247.5	15.3	981	4	US-09-902-540-2515	Sequence 2515, Ap
c	247.5	15.3	981	4	US-09-902-540-2515	Sequence 2515, Ap
42	247	15.3	14367	4	US-09-902-540-1113	Sequence 1113, Ap
c	247	15.3	14367	4	US-09-902-540-1113	Sequence 1113, Ap
44	245.5	15.2	1062	4	US-09-328-352-2595	Sequence 2595, Ap
c	245.5	15.2	1062	4	US-09-328-352-2595	Sequence 2595, Ap
46	245.5	15.2	945	3	US-09-134-001C-2568	Sequence 2568, Ap
c	245.5	15.2	945	3	US-09-134-001C-2568	Sequence 2568, Ap
48	244.5	15.1	1266	4	US-09-252-991A-3169	Sequence 3169, Ap
c	244.5	15.1	1266	4	US-09-252-991A-3169	Sequence 3169, Ap
50	241	14.9	1056	4	US-09-252-991A-2799	Sequence 2799, Ap
c	241	14.9	1056	4	US-09-252-991A-2799	Sequence 2799, Ap
52	241	14.9	1026	4	US-09-902-540-3767	Sequence 3767, Ap
c	241	14.9	1026	4	US-09-902-540-3767	Sequence 3767, Ap
54	237.5	14.7	984	4	US-09-902-540-1183	Sequence 1183, Ap
c	237.5	14.7	984	4	US-09-902-540-1183	Sequence 1183, Ap
56	237.5	14.7	41310	4	US-09-902-540-1264	Sequence 1264, Ap

ALIGNMENTS

RESULT 1

US-09-630-983A-12
; Sequence 12, Application US/09630983A
; Patent No. 6630330
; GENERAL INFORMATION:
; APPLICANT: Porro, Michael
; TITLE OF INVENTION: Ascorbic Acid Production from Yeast
; FILE REFERENCE: 2028.594000
; CURRENT APPLICATION NUMBER: US/09/630,983A
; CURRENT FILING DATE: 2000-08-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-630-983A-12

Alignment Scores:
Pred. No.: 3.84e-190 Length: 960
Score: 1619.00 Matches: 319
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-606-300-11 (1-319) x US-09-630-983A-12 (1-960)

Qy	1	MetThrLysLleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGly	20
Db	1	ATGACGAATAATAGCTTCGAGCTTTGGGGAACACAGGGCTTAAAGTTAGCGCGTGGT	60
Qy	21	PheGlyAlaSerProLeuGlySerValPheClyProValAlaGluAspAlaValala	40
Db	61	TTTGGTGCCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGGAAGATGATCGCGTCGCC	120


```
US-10-606-300-11 (1-319) x US-07-855-793-3 (1-2682)
QY 5 GluLeuArgAlaLeuGlyAsnThrGlyLeuValSerAlaValGlyPheGlyAlaSer 24
DB 850 GAACCGCCACCATCCCGGACGAGAGCTCTGTCGGCTGCTCGGCTTCGGCGGGCC 909
QY 25 ProLeuGlySerValPheGlyProValAlaGluAspAlaValAlaThrValArgGlu 44
DB 910 CCCATCGGCAACCTTACCGGCGAGTCCCGGAGCAGGAAGCGTGCAGCGCGT 969
QY 45 AlaPheArgLeuGlyIleAsnPheAspThrSerProTyrTyrGlyThrLeuSer 64
DB 970 GCCTGGGAGGCGGCTCGCTATTTCATACCGCGCGCACTACGGGCTGGGCTCTCG 1029
QY 65 GluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAspTyrIleVal 84
DB 1030 GAACGCCCATGGGCGCGCTCCAGGGCAGGAC-----CGGACAGTTACGTGCTC 1083
QY 85 AlaThrLysCysGlyArgTyr-----LysGlu 93
DB 1084 AGCAGAAAGATCGGGCGGCTCTGCGGCCCAACCGCTCCCGGAGGCAAGCAGCACCGAG 1143
QY 94 GlyPhe-----AspPheSerAlaGluArgValArg 103
DB 1144 GCCTTCGACGTCCTCCGACGACTCTATCCGGGTCCGCGACTACTCCCGACGGGGTGCTG 1203
QY 104 LysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAlaPheIleLeuHisCys 123
DB 1204 CGCTCCATCGAGGAAGCTCGAGCGGCTGGGACCGACCGGATCGCATGTCTACATC 1263
QY 124 HisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIleProAlaLeuGln 143
DB 1264 CACGACCTGAC---GACTACTGGACCGCGCGTGGAGGCGCGCCCGCGCTGTGCC 1320
QY 144 LysLeuLysGlnGluGlyThrArgPheIleGlyIle-----ThrGlyLeu 159
DB 1321 GCCCTGGCGGACGAAGGGGTCTACGAGCGCTGGGCGCGAGCATGAACCAATCCGGGATG 1380
QY 160 ProLeuAspIlePheThrTyrValLeuAspArgValProGlyThrValAspValIle 179
DB 1381 ---CTCCACAGTTTCATCCGAAACGGAC-----ATTGACGTGTC 1419
QY 180 LeuSerTyrCysHisTyrGlyValAsnAspSerThrLeuLeu----- 193
DB 1420 ATGCTGGCGCGCGGTAC-----ACGCTCTGGAGCAGGGTGCAGCACAG 1464
QY 194 AspLeuLeuProTyrLeuLysSerLysGlyValGlyValIleSerAlaSerProLeuAla 213
DB 1465 GACCTGCTTCTGCTGCTGCTGAGCGCGGGTGGCGTGGTGAACGTCGCTCTCAAT 1524
QY 214 MetGlyLeuLeuThrGluGlnGlyProPro-----GluTyrHisProAla 228
DB 1525 TCGGCGCTTGTGTCCAGAGCGGCGCGCGGCAAGCCACCTACACTACGCGCGCGCC 1584
QY 229 SerProGluLeuLysSerAlaSerLysAlaValAlaHisCysLysSerLysGlyLys 248
DB 1585 CGCGAGGAATGCTCGACAGGCGCAACCTGTGGCAGACGCTCGGAATCCACGGCACC 1644
QY 249 LysIleThrLysLeuAlaLeuGlnTyrSerLeuAlaAsnLysGluIleSerSerValLeu 268
DB 1645 ACCTGCGCGCGCGCGCTCCACTTCCCTTACCAGACACCCCGCGCTCACCAGCGTGTGC 1704
QY 269 ValGlyMetSerSerValSerGlnValGluGluAsn-----ValAlaAlaValThrGlu 286
DB 1705 CTTGGCATCGGACCGCGCGCAGGTGAAGCAAAACCTGGACCTCGCCTCGACAGCGTCTC 1764
QY 287 LeuGluSerLeuGlyMetAsp 293
DB 1765 CCGGACACAGCTGTGGGCTGAC 1785
RESULT 3
US-09-252-991A-10047
; Sequence 10047, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10047
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10047
Alignment Scores:
Pred. No.: 5,28e-29 Length: 1104
Score: 317.00 Matches: 84
Percent Similarity: 50.17% Conservative: 60
Best Local Similarity: 29.27% Mismatches: 103
Query Match: 19.58% Indels: 40
DB: 4 Gaps: 8
US-10-606-300-11 (1-319) x US-09-252-991A-10047 (1-1104)
QY 19 ValGlyPheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAla 38
DB 154 ATCGGTCTCGGCGCGCGACCGCTCGGCAACATGTTCCATCGCTCAGCAAGAACCCGCC 213
QY 39 ValAlaThrValArgGluAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyr 58
DB 214 GACGCCACCTCAATGCGCTCGGACCGCGCTTTCGCTATTACGAGCTGTCCCGCGAC 273
QY 59 TyrGlyGlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValPro 78
DB 274 TACGCGCGGCGCTCGCGGACGCGCTTCGGGCGCTGTGCTGAGTGGC-----AAGCG 327
QY 79 ArgSerAspTyrIleValAlaThrLysCysGlyArgTyr----- 91
DB 328 CGTGACGAGTACGTCTGAGCACCAAGGTCCGGCGCTGTCTGCAGCCCGCGCAGCCAGCG 387
QY 92 -----LysGluGlyPheAspPheSerAla 99
DB 388 GAAACGCCAAGCCCTCTGTCGAGAGTGTGCCCAACAGCGTGTGCCGATATTTCGGCC 447
QY 100 GluArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAsp 119
DB 448 GACGCGCACCGCTCCATCGAGGACAGCTTGAGCGCATGGGCGTGGCGTGCATCGCTCAT 507
QY 120 IleLeuHisCysHisAspIle-----GluPheGlySer-----Leu 131
DB 508 GTGTGTTCATCCACGAGCTTTCGGAAGACCAATGGGCGCCGCAATGGCGGAGTACTTC 567
QY 132 AspGlnIleValSerGluThrIleProAlaLeuGlnLysLeuGlnGluGlyLysThr 151
DB 568 CAGCAGGCGCATGAAGCGCGCGGAGCGCTGACAGCTGCGGATGAAGGCGTGTATT 627
QY 152 ArgPheIleGlyIleThrGlyLeuProLeuAspIlePheThrTyrValLeuAspArgVal 171
DB 628 CGCGCTGGGCGCTGGGGGTCACTGTGTGGAGCATGCCCGCTGGCTCTGGAGCAGAGC 687
QY 172 ProProGlyThrValAspValIleLeuSerTyrCysHisTyrGlyVal-----AsnAsp 189
DB 688 GACCG-----AACGTGCTCTGTCGCGCGCGCTACTCGCTCTGGAGCAGCAGC 738
QY 190 SerThrLeuLeuAspLeuLeuProTyrLeuLysSerLysGlyValGlyValIleSerAla 209
DB 739 GAGCGCTGGACACCTCTTCCCGACCTGCGCAGGACGCGCTGTGGCTGTGCTTGGC 798
```

QY 210 SerProLeuAlaMetGlyLeuLeuThrGluGlnGlyProProGluTrpHisProAlaSer 229
Db 799 GGGCGGTTCAACTCCGGGCTCTGGCGGTGGCGATCACTAGTACGACCATCCCA 858
QY 230 ProGluLeu-----LysSerAlaSerLysAlaAlaValAlaHisCysLysSerLys 246
Db 859 CCACAGGTGCGCCACGCGAGGAGCAGCTCAAGGCGCTGCGGACACTGC----- 909
QY 247 GlyLysLysIleThrLysLeuAlaLeuGlnTyrSerLeuAlaAsnLysGluIleSerSer 266
Db 910 GGGCTGCATCTGGGCGCGCGCTGCTGCTTCTGCTGCGCAACCCGGTGGTGGCTCG 969
QY 267 ValLeuValGlyMetSerSer 273
Db 970 GTCATCCGGGCGACGCCAAT 990

RESULT 4

US-09-252-991A-9791/c
; Sequence 9791, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9791
; LENGTH: 2514
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9791

Alignment Scores:
Pred. No.: 2,098-28 Length: 2514
Score: 317.00 Matches: 84
Percent Similarity: 50.1% Conservative: 60
Best Local Similarity: 29.2% Mismatches: 103
Query Match: 19.5% Indels: 40
Gaps: 4

US-10-606-300-11 (1-319) x US-09-252-991A-9791 (1-2514)

QY 19 ValGlyPheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAla 38
Db 1029 ATCGGTCTCGGGGCGGCGACCGCTGGCGCAATGTTCCATCGCTCAGCGAAGAACCGCC 970
QY 39 ValAlaThrValArgGluAlaPheArgLeuGlyIleAsnPheAspThrSerProTyr 58
Db 969 GACGCGACCCCTCAATCCCGCTGGGACCGCGCTTTCGCTATTACGAGCTGTCGCGGCAC 910
QY 59 TyrGlyGlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValPro 78
Db 909 TACGCGCGGGGCTCGCGAGCAGCGCTTCGGGCGCTGCTGAGTGGC-----AAGCG 856
QY 79 ArgSerAspTyrIleValAlaThrLysCysGlyArgTyr----- 91
Db 855 COTGACGAGTACGCTCTGAGCACCAAGGTGCGGCGCTGCTGACGCGCGCCAGCCAGCGCG 796
QY 92 -----LysGluGlyPheAspPheSerAla 99
Db 795 GAAACGCCAACCCCTTCGTGACGAGCTGCCCAACAGCGGTGCCGAGTATTTCGCGC 736
QY 100 GluArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAsp 119
Db 735 GACGCGGACCGCGCTTCCATCGAGACAGCTGGAGCGCATGGGCGTGCATCGCTCGAT 676

QY 120 IleLeuHisCysHisAspIle-----GluPheGlySer-----Leu 131
Db 675 GTGTGTTTCATCCAGCGTTTCCGAAGACCAATGGGCGCCGCAATGGCGGAGTACTTTC 616
QY 132 AspGlnIleValSerGluThrIleProAlaLeuGlnLysLeuLysGlnGluGlyLysThr 151
Db 615 CAGCAGGCCATGACGCGCGCGGAGCGCTGACCCAGCTGCGCGATGAGCGGTGATT 556
QY 152 ArgPheIleGlyIleThrGlyLeuProLeuAspIlePheThrTyrValLeuAspArgVal 171
Db 555 CGCGCTGGGCGCTGGGGGTCAACTGTGTGAGCATGCGCGCTGGCTCTGGAGCAGAGC 496
QY 172 ProProGlyThrValAspValIleLeuSerTyrCysHisTyrGlyVal-----AsnAsp 189
Db 495 GACCCG-----AAGCTGTTCTGTGCGCGCGCTACTCGCTCTGGAGCAGCAGC 445
QY 190 SerThrLeuLeuAspLeuProTyrLeuLysSerLysGlyValGlyValIleSerAla 209
Db 444 GAGCGCTGGACACCTGTTCCGACCTGCCAGGACGCGAGCTTGGGTGCTGCTGCTGGC 385
QY 210 SerProLeuAlaMetGlyLeuLeuThrGluGlnGlyProProGluTrpHisProAlaSer 229
Db 384 GGGCGGTTCAACTCCGGGCTCTGGCGGTGGCGATCACTAGTACGACCATCCCA 325
QY 230 ProGluLeu-----LysSerAlaSerLysAlaAlaValAlaHisCysLysSerLys 246
Db 324 CCACAGTCCGCCAGCGCAGGAGCAGCTCAAGCGCGCTGCCGAGCACTGC----- 274
QY 247 GlyLysLysIleThrLysLeuAlaLeuGlnTyrSerLeuAlaAsnLysGluIleSerSer 266
Db 273 GGGCTGCATCTGGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 214
QY 267 ValLeuValGlyMetSerSer 273
Db 213 GTCATCCGGGCGACGCCAAT 193

RESULT 5

US-09-902-540-3178
; Sequence 3178, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 3178
; LENGTH: 1044
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-3178

Alignment Scores:
Pred. No.: 9,8e-29 Length: 1044
Score: 314.50 Matches: 87
Percent Similarity: 46.13% Conservative: 62
Best Local Similarity: 26.93% Mismatches: 119
Query Match: 19.43% Indels: 55
Gaps: 9

US-10-606-300-11 (1-319) x US-09-902-540-3178 (1-1044)

QY 4 IleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGlyPheGlyAla 23
Db 34 GTTCCCAACGCGCGCGGCAACAAAGTTCCGCGCCACCAACCGCATTTGGCTGGCGCGC 93
QY 24 SerProLeuGlySerValPheGlyProValAlaGluAspAlaValAlaThrValArg 43

MEDIUM TYPE: CD/ROM ISO9660
 COMPUTER: PC
 OPERATING SYSTEM: <Unknown>
 SOFTWARE: ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/107,532A
 FILING DATE: 30-Jun-1998
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/085,598
 FILING DATE: 14 May 1998
 APPLICATION NUMBER: 60/051571
 FILING DATE: July 2, 1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Ariniello, Pamela Denek
 REGISTRATION NUMBER: 40,489
 REFERENCE/DOCKET NUMBER: GTC-012
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (781)893-5007
 TELEFAX: (781)893-8277
 INFORMATION FOR SEQ ID NO: 3237:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 999 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: circular
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Enterococcus faecium
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (B) LOCATION 1...999
 SEQUENCE DESCRIPTION: SEQ ID NO: 3237:

US-09-107-532A-3237

Alignment Scores:
 Pred. No.: 1.53e-26 Length: 999
 Score: 296.50 Matches: 92
 Percent Similarity: 47.98% Conservative: 62
 Best Local Similarity: 28.66% Mismatches: 129
 Query Match: 18.31% Indels: 39
 DB: 4 Gaps: 10

US-10-606-300-11 (1-319) x US-09-107-532A-3237 (1-999)

Qy 5 GluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGlyPheGlyAlaSer 24
 Db 58 GAAATAAGA---ATCGACATTCACAGTCTACGCAGAACAGCTTGGATTAGGACAAAT 114
 Qy 25 ProLeuGly-----SerValPheGlyProValAlaGluAspAlaValAlaThrVal 42
 Db 115 GCCGTGGAGGACATAATCTTTTGGCCGTTTGGAGATGAGACAGGAAACAAAGTAGTT 174
 Qy 43 ArgGluAlaPheArgLeuGlyIleAsnPheAspThrSerProTyrTyrGlyGlyThr 62
 Db 175 GTACAGCTTTGAATAGCGGATCAATTTGATCGATCCGCTTATGCTTACGGAACGGC 234
 Qy 63 LeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAspTyr 82
 Db 235 CGTTCCGAAGAACTGATTGGCGAGGTCTGAAGGAAAGAAATATGATCGCTCAGGTGC 294
 Qy 83 IleValAlaThrLysCysGlyArgTyrLys-----GluGlyPheAspPheSerAla 99
 Db 295 GTCATCGCTCAAAAGCTGCTCATGTACCAATAAAGGCGAACATTTGATAATTCCTCT 354
 Qy 100 GluArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAsp 119
 Db 355 GAGTTTCTCAACAATCGGTGAAGATGATCACTAAACGCTGCAACTGATTATCATCGAC 414
 Qy 120 IleLeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIle 139
 Db 415 ATTTTATATATCCAC-----TTTCCGGATGAAGACACACACCAACAAATAATGAATCAGTT 465

Qy 140 ProAlaLeuGlnLysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGlyLeu 159
 Db 466 GCTACATTGACGAAATTAAGAAAGACGAGAAATCCGAGCAGTCGGTGTGTCAAACTTC 525
 Qy 160 ProLeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAspValIle 179
 Db 526 ACGTTAGAA-----CAATTGAAAGAACGGAATGCAGATGATATGTCGATGTAGTA 576
 Qy 180 -----LeuSerTyrCysHisTyrGlyValaAspSerThrLeuLeuAspLeuLeu 196
 Db 577 GAAGATAAATACAGTTTGTATCCATCGGACGCTGAAAAA-----GAATTATTC 624
 Qy 197 ProTyrLeuLysSerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeu 216
 Db 625 CCTTATCTTGAAAAAACAAATCAGTTTGTTCCTTATTTCTCTTCTTCCGGTTG 684
 Qy 217 LeuThr-----GluGlnGlyProProGlu 224
 Db 685 CTGACAGGAAATATGAATTTGGTGAGGAAAAACAATTTGGTGAGGAGATCCTCGAAAA 744
 Qy 225 TrpHisPro-----AlaSerProGluLeuLysSerAlaSerLysAlaAla 239
 Db 745 AGAATCCAGATTTCACAGAGAACGTTTCAGAGAAATTTTAAACCGCTGT-CGATGTACT 803
 Qy 240 ValAlaHisCys-LysSerLysGlyLysLysIleThrLysLeuAlaLeuGlnTyrSerLe 259
 Db 804 TCGCCCATTCGTAAGATACCAAGCACCTCCTCGCGCAACTTGTTTTAGCTTGTGTACAT 863
 Qy 259 uAlaAsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGlu 279
 Db 864 GAAGATCCACGAGTGTCTGTAGTTATACCTGGAGCTAAAGCGTCCAGAGCAAGTATCTGA 923
 Qy 279 uAsnValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerG 299
 Db 924 CAATGTCCAGCTTTA---GATTTCATTTTATCAATGAAGATTATCAACCAATTGATGA 980
 Qy 299 u 299
 Db 981 A 981

RESULT 12

US-09-328-352-1953
 ; Sequence 1953, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 1953
 ; LENGTH: 1005
 ; TYPE: DNA
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-1953

Alignment Scores:
 Pred. No.: 1.79e-26 Length: 1005
 Score: 296.00 Matches: 93
 Percent Similarity: 44.93% Conservative: 62
 Best Local Similarity: 26.96% Mismatches: 128
 Query Match: 18.28% Indels: 62
 DB: 4 Gaps: 10

US-10-606-300-11 (1-319) x US-09-328-352-1953 (1-1005)

Qy 3 LysIleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGlyPheGly 22
 Db 1 AAAATGAATATAGAACTTTAGGCAAAACAGCGCGAAAGATATCAGCACCTTGGTTAGGT 60

Qy	23	AlaSerProLeuGlySerValPheGlyProValAlaGluAspAspAlaValAlaThrVal	42
Db	61	TGCATGGGAATGAGCTTTGGCTATGGTCCTCAGACGACACACAAAGAACTTGAACCTCTA	120
Qy	43	ArgGluAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyrTyrGlyGlyThr	62
Db	121	GAAGAAGCCTTAGACCTCGCATTAACCTTTGGGACACTGCCGACATGCTACGGCAACGGC	180
Qy	63	LeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAspTyr	82
Db	181	GCAACCAAGTTTGTCTCA-----AAAGTTTTAGAAAAACATCGTGCAAAAGT	231
Qy	83	IleValAlaThrLysCysGly---ArgTyrLysGluGly-----	94
Db	232	TTCTTGGCACCAAATTTGGCTTCGCCTATAAAGAGATAACCTTAACCGCAAAAACCTCT	291
Qy	95	-----PheAspPheSerAlaGluArgValArgLysSerIleAspGluSerLeu	110
Db	292	TTAGAGCTCTATATCGACGGCTCGCTGAGTGGATTAAAGTTGGCGTTCGAAAACAGCTTA	351
Qy	111	GluArgLeuGlnLeuAspTyrValAspIleLeuHisCysHisAspIleGluPheGlySer	130
Db	352	AGACGCTTAAATACAGACGTGATTGATTGATTATGTCGCACCGTATAGAC-----	402
Qy	131	LeuAspGlnIleValSerGluThrIleProAlaLeuGlnLysLeuLysGlnGluGlyLys	150
Db	403	CCGAATGTTCAGTTGAAGATACCAATTCGGCGAATGGCAGACTTGGTCAACAAAGGTAA	462
Qy	151	ThrArgPheIleGlyIleThrGlyLeuProLeuAspIlePheThrTyrValLeuAspArg	170
Db	463	GTTCGTTATTAGGGCTGAGCGAAGCGTCGGCTGAACTATTTCGTAAAGCACATCGCAT	522
Qy	171	ValProProGlyThrValAspValIleLeuSerTyrCysHisTyrGlyValAsnAspSer	190
Db	523	CACCCGATTTCGACGCGTTCAA-----LeuProTyrLeuLysSerLysGlyVal	204
Qy	191	ThrIleIleLeuAspLeu-----LeuProTyrLeuLysSerLysGlyVal	204
Db	553	TCGCTACTTACCGTGAGTTTGAACAAACCCATTTCMAACCATTCGTGAGTAGGTATT	612
Qy	205	GlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGlu-----	219
Db	613	AGCCTTGGCATATTCTCCGTTATCCGCTGTTTAAATTACCAATACGTTGATGTAAAT	672
Qy	220	-----GlnGlyProProGluTyr-----	225
Db	673	AACTAGATGAAATGATTTCGTCGCCGATTAACCGTTATACAGGC-----GATAACTGG	729
Qy	226	HisProAlaSerProGluLeuLysSerAlaSerLysAlaAlaValAlaHisCysLysSer	245
Db	730	-----AAAAATAACCAAGCTTGGCGCAAGCATTTAGTGAATTTGCTCAAGC	777
Qy	246	LysGlyLysLysIleThrLysLeuAlaLeuGlnTyrSerLeuAlaAsnLysGluIleSer	265
Db	778	AAAAATGCGACACGCCCTCAACTGGCTTTGGCTTGGATTTTGGCGCAAGGTGATGACATT	837
Qy	266	SerValLeuValGlyMetSerSerValSerGlnValGluGluAsnValAlaAlaValThr	285
Db	838	ATTCCAATTCGCGGTACGCGTAAATATCAGCGCACTTGTGAAATTCGGGTGCTGTA---	894
Qy	286	GluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluValGluAlaIleLeuGlu	305
Db	895	-----GACCTTCATTACACGGCGCAGATTGGCAGAAATCGACGGGATTTTGGC	945
Qy	306	ProValLysAsnLeu	310
Db	946	AGATACCAACATG	960

RESULT 13
US-09-134-001C-289
; Sequence 289, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:

Pred. No.: 7, 91e-26 Length: 1044
Score: 291.00 Matches: 94
Percent Similarity: 46.53% Conservative: 60
Best Local Similarity: 28.40% Mismatches: 135
Query Match: 17.97% Indels: 42
DB: 4 Gaps: 11

US-10-606-300-11 (1-319) x US-09-673-198-9 (1-1044)

```
QY 4 IleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSer-----Ala 18
   ::::: ||||| ||| |||||
Db 73 ATGCAATACAACCCCTTAGGAAAAACGACCTTCGCGTTTCGCGACTTTCGCGCTGT 132
QY 19 ValGlyPheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAla 38
   ::||| ||||| |||||
Db 133 ATGACCTTTTGGCGAGCCAGATCGCGGTAAATCACGCATGACACCTCCCGGAAAGACAGC 192
QY 39 ValAlaThrValArgGluAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyr 58
   ::::: ||||| ||||| ||||| ||||| ||||| |||||
Db 193 CGTCCCATAAATTAACGTCACCTGGAAGCGGCATAAATTTCTTTGATACCGCCACAGT 252
QY 59 TyrGlyGlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValPro 78
   ||||| ||||| ::::: |||||
Db 253 TATTCTGACGCGCAGCAGCAGAGATCGTCGTCGCGCAGCTCGCGGATTTC---GCCCGT 309
QY 79 ArgSerAspTyrIleValAlaThrLys-----CysGlyArgTyrLysGluGly 94
   ||||| ||||| ::||| ||||| |||||
Db 310 CGTGAAGACGTCGTCGTTGCGACCAAAAGTGTTCATCGCGTTGGTGTATTTACCGGAAGGA 369
QY 95 PheAspPheSerAlaGluArgValArgLysSerIleAspGluSerLeuGluArgLeuGln 114
   ||||| ||||| ::::: ||||| ||||| ||||| |||||
Db 370 -----TTATCCCGTCGCAAAATTTTGGCTCTATCGACGACAGCTCGAGCTCTCGGC 423
QY 115 LeuAspTyrValAspIleLeuHisCysHisAspIleGluPheGlySerLeuAspGlnIle 134
   ::||| ||||| ||||| ||||| ::::: |||||
Db 424 ATGGATTATGCGATATCTGCAAAATTCATCGCTGGATTACAAACAG-----CCG 474
QY 135 ValSerGluThrIleProAlaLeuGlnLysLeuLysGlnGluGlyLysThrArgPheIle 154
   ::||| ||||| ||||| ::||| ||||| |||||
Db 475 ATCGAAGACGCTGGAAGCCCTCAACGACGTGTGTAAAGCCGGGAAAGCGCGTTATATC 534
QY 155 GlyIleThrGlyLeuProLeuAspIlePheThrTyrValLeuAsp---ArgValProPro 173
   ||||| ::||| ||||| ||||| ||||| ::|||
Db 535 GCGCGCTCATCAATCAGCGCTTCGAGTTTGTCTAGGCGACTGGAATCTCAAAAACAGCAC 594
QY 174 GlyThrValAspValIleLeuSerTyrCysHisTyrGly---ValAsnAspSerThrLeu 192
   ||||| ::||| ||||| |||||
Db 595 GCGTGGCGCGAGTTGTGATGATGAGGATCACTACATCTGATTTATCGTGAAGAGAG 654
QY 193 LeuAspLeuLeuProTyrLeuLysSerLysGlyValGlyValIleSerAlaSerProLeu 212
   ::||| ||||| ||||| ||||| |||||
Db 655 CGCGAGATGTACCACTGTGTTATCAGGAGGCGTGGCGGTAAATTCATGAGCCCGCTG 714
QY 213 AlaMetGlyLeuThrGluGlnGlyProProGluTyrHisProAlaSerProGluLeu 232
   ||||| ||||| ||||| |||||
Db 715 GCAAGGGGCGCTGACGCGT-----CCGTGGGGAGAAACTACCGCACGACTG 762
QY 233 LysSer-----AlaSerLysAlaAlaVal 240
   ||||| |||||
Db 763 GTGTCGTGATGAGTGGGNAANAATCTCTATAAGAAAGCGATGAAATCAGCGCGAGATC 822
QY 241 AlaHis-----CysLysSerLysGlyLysIleThrLysLeuAlaLeu 255
   ||||| ::||| |||||
Db 823 GCAGAGCGGTTAACAGCGCTCAGTGAAGAACTGGGGGCGACACGAGCACAAAGTTGCGCTG 882
QY 256 GlnTyrSerLeuAlaAsnLysGluIleSerSerValLeuValGlyMetSerSerValSer 275
   ::||| ||||| ||||| ||||| |||||
Db 883 GCTGTGTTGAGTAAACCGGGCATTCGCCGCCGATTATCGGAATTCGCGCGAAGAA 942
QY 276 GlnValGluGluAsnValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGlu 295
   ||||| ||||| ||||| |||||
Db 943 CAGCTTGATGACTATTGACGCGGTG-----GATATCACTTTGAAGCGGAA 990
```

QY 296 ThrLeuSerGluValGluAlaIleLeuGluPro 306
Db 991 CAGATTGCCGAACCTGGAAACCGCGGTATTAACCG 1023

Search completed: November 13, 2005, 11:18:40
Job time : 231 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 13, 2005, 08:41:00 ; Search time 894 Seconds
(without alignments)
2950.864 Million cell updates/sec

Title: US-10-606-300-11
Perfect score: 1619
Sequence: 1 MTKIELRALGNTGLKLSAVG.....VEAILEPVONLTWPSGIHQN 319

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0

Maximum Match 0%

Post-processing: Minimum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_gpool/h/US10606300/runat_07112005_094939_8214/app query.fasta_1.455
-DB=Published Applications NA -OFT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10606300@cgn1.1.480@runat_07112005_094939_8214
-NCPU=6 -ICPU=3 -NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*
23: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
24: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
25: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq.*
26: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
27: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
28: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1619	100.0	960	19	US-10-240-136A-4
2	1619	100.0	960	22	US-10-606-300-12
3	1279	79.0	1361	19	US-10-424-599-119463
4	1213	74.9	1616	20	US-10-437-963-11457
5	1208.5	74.6	1715	21	US-10-425-115-95235
6	1199.5	74.1	1434	20	US-10-767-701-13219
7	1191.5	73.6	1295	21	US-10-425-115-95234
8	908	56.1	1134	19	US-10-425-114-31543
9	779.5	48.1	564	20	US-10-021-323-3555
10	593.5	36.7	1038	26	US-11-097-143-28022
11	546.5	33.8	3724	26	US-11-097-143-28021
12	544.5	33.6	1051	26	US-11-097-143-40478
13	517	31.9	3308	26	US-11-097-143-40477
14	497	30.7	563	20	US-10-021-323-3629
15	458	28.3	3014	26	US-11-097-143-7144
16	420.5	26.0	885	26	US-11-097-143-7145
17	389.5	24.1	9025608	16	US-10-156-761-1
18	385	23.8	984	16	US-10-156-761-1393
19	368.5	22.8	984	16	US-10-156-761-7263
20	368.5	22.8	9025608	16	US-10-156-761-1
21	342	21.1	10400	26	US-11-097-143-27337
22	342	21.1	10418	26	US-11-097-143-40462
23	342	21.1	10451	26	US-11-097-143-26110
24	337.5	20.8	921	18	US-10-369-493-46874
25	335	20.7	2838	20	US-10-437-963-68518
26	329.5	20.4	912	18	US-10-282-122A-9381
27	329	20.3	1246	19	US-10-425-114-5787
28	329	20.3	1296	21	US-10-425-115-172723
29	329	20.3	1544	21	US-10-739-930-2881
30	327	20.2	975	18	US-10-369-493-28562
31	322	19.9	924	18	US-10-369-493-31322
32	322	19.9	987	18	US-10-282-122A-21647
33	316	19.5	921	18	US-10-369-493-40147
34	316	19.5	1038	18	US-10-282-122A-19546
35	314	19.4	2256646	20	US-10-470-565-1
36	313	19.3	981	9	US-09-815-242-6089
37	313	19.3	981	18	US-10-369-493-24549
38	313	19.3	981	18	US-10-282-122A-7060
39	310.5	19.2	906	18	US-10-282-122A-35420
40	310.5	19.2	993	16	US-10-156-761-1151
41	310	19.1	933	18	US-10-282-122A-9361
42	308	19.0	1791	19	US-10-424-599-13910
43	307.5	19.0	933	18	US-10-369-493-43066
44	306.5	18.9	999	18	US-10-282-122A-38764
45	305.5	18.9	972	9	US-09-815-242-7749

ALIGNMENTS

RESULT 1

US-10-240-136A-4
; Sequence 4, Application US/10240136A
; Publication No. US2004005323A1
; GENERAL INFORMATION:
; APPLICANT: Ascortex Limited
; TITLE OF INVENTION: Gene Sequence
; FILE REFERENCE: JPD/P100606US-PCT
; CURRENT APPLICATION NUMBER: US/10/240.136A
; CURRENT FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: GB 0007651.3
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1

```
; SEQ ID NO 4
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(960)
; OTHER INFORMATION:
US-10-240-136A-4

Alignment Scores:
Pred. No.: 1.33e-172 Length: 960
Score: 1619.00 Matches: 319
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 19 Gaps: 0

US-10-606-300-11 (1-319) x US-10-240-136A-4 (1-960)
QY 1 MetThrLysIleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGly 20
DB 1 ATGCGAAAATAGAGCTTCGAGCTTTGGGGAACACAGGCTTAAGGTTAGCGCGCTTGGT 60
QY 21 PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAla 40
DB 61 TTTGGTGCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCCGAAGATGATGCCGTGCC 120
QY 41 ThrValArgGluAlaPheArgLeuGlyIleAsnPheAspThrSerProTyrTyrGly 60
DB 121 ACCGTGGCGAGGCTTCGCTCGGTATCAACTTTCGACACCTCCCGGTATTATGGA 180
QY 61 GlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSer 80
DB 181 GGAACACATGCTCGAGAAATGCTTGGTAAGGACTAAAGGCTTTGCAAGTCCCTAGAAGT 240
QY 81 AspTyrIleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerIleGlu 100
DB 241 GACTACATTTGGTGTACTAAGTGTGGTAGATATAAGAGGTTTGTGATTTTCAGTGTCTG 300
QY 101 ArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 120
DB 301 AGAGTAAGAAAGATGATTGACGAGCTTTGGAGAGCTTCAGCTGATTTGTTGACATA 360
QY 121 LeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePro 140
DB 361 CTTTCAATTTGCGATGATGATTCGGGTCTCTTGATCAGATTGTGAGTGAACAATTCCT 420
QY 141 AlaLeuGlnLysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGlyLeuPro 160
DB 421 GCTCTTCAGAACTGAAACAAAGGGAAGATCACAAGTTAGTCTTGCATATACAGTTTAGCA 480
QY 261 AsnLysGluIleSerValLeuValGlyMetSerSerValSerGlnValGluGluAsn 280

; SEQ ID NO 4
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(960)
; OTHER INFORMATION:
US-10-240-136A-4

Alignment Scores:
Pred. No.: 1.33e-172 Length: 960
Score: 1619.00 Matches: 319
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 19 Gaps: 0

US-10-606-300-12
; Sequence 12, Application US/10606300
; Publication No. US20050019879A1
; GENERAL INFORMATION:
; APPLICANT: Sauer, Danilo
; TITLE OF INVENTION: Ascorbic Acid Production from Yeast
; FILE REFERENCE: 2028.594000
; CURRENT APPLICATION NUMBER: US/10/606.300
; CURRENT FILING DATE: 2003-06-25
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 12
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-606-300-12

Alignment Scores:
Pred. No.: 1.33e-172 Length: 960
Score: 1619.00 Matches: 319
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-10-606-300-11 (1-319) x US-10-606-300-12 (1-960)
QY 1 MetThrLysIleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGly 20
DB 1 ATGCGAAAATAGAGCTTCGAGCTTTGGGGAACACAGGCTTAAGGTTAGCGCGCTTGGT 60
QY 21 PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAla 40
DB 61 TTTGGTGCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCCGAAGATGATGCCGTGCC 120
QY 41 ThrValArgGluAlaPheArgLeuGlyIleAsnPheAspThrSerProTyrTyrGly 60
DB 121 ACCGTGGCGAGGCTTCGCTCGGTATCAACTTTCGACACCTCCCGGTATTATGGA 180
QY 61 GlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSer 80
DB 181 GGAACACATGCTCGAGAAATGCTTGGTAAGGACTAAAGGCTTTGCAAGTCCCTAGAAGT 240
QY 81 AspTyrIleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerIleGlu 100
DB 241 GACTACATTTGGTGTACTAAGTGTGGTAGATATAAGAGGTTTGTGATTTTCAGTGTCTG 300
QY 101 ArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 120
DB 301 AGAGTAAGAAAGATGATTGACGAGCTTTGGAGAGCTTCAGCTGATTTGTTGACATA 360
QY 121 LeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePro 140
DB 361 CTTTCAATTTGCGATGATGATTCGGGTCTCTTGATCAGATTGTGAGTGAACAATTCCT 420
QY 141 AlaLeuGlnLysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGlyLeuPro 160
DB 421 GCTCTTCAGAACTGAAACAAAGGGAAGATCACAAGTTAGTCTTGCATATACAGTTTAGCA 480
QY 161 LeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeu 180
```

Db 481 TTAGATATTTTCACTATGTTCTTGTATCGATGCTCCAGGAGCTGTGATGATTTG 540
Qy 181 SerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLys 200
Db 541 TCATACCTGTCATACGGGCTTAATGATTCGAGCTTGCTGGATTTACTACCTACTTGTAG 600
Qy 201 SerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGln 220
Db 601 AGCAAGGCTGGGTGTGTAAGTCTCTCCATTTAGCAATGGGCTCTCTTACAGAACAA 660
Qy 221 GlyProProGluTyrHisProAlaSerProGluLeuLysSerAlaSerLysAlaVal 240
Db 661 GGTCTCTGTAATGGACCT 720
Qy 241 AlaHisCysLysSerLysGlyLysIleThrLysLeuAlaLeuGlnTyrSerLeuAla 260
Db 721 GCTCACTGCAATCAAAGGCGCAAGATCACAAAGTTAGCTCTGCAATACAGTTAGCA 780
Qy 261 AsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGln 280
Db 781 AACAAAGGAGATTTCTGCTGGTGTGGTGGATGAGCTCTGTCTCACAGGTAGAAAGAAAT 840
Qy 281 ValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnThrLeuSerGluVal 300
Db 841 GTTCAGCAGATTACAGAGCTTGAAGTCTGGGATGGATCAAGAAACTCTGTCTGAGGTT 900
Qy 301 GluAlaIleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGlnAsn 319
Db 901 GAAGCTATTCGAGCCTGTAAGANCTGATGCGCCAGTGGATCCATCCATCAGAAC 957

RESULT 3

US-10-424-599-119463

; Sequence 119463, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 119463

; LENGTH: 1361

; TYPE: DNA

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT3847_78884C.1

US-10-424-599-119463

Alignment Scores:

Pred. No.: 6,246-134 Length: 1361

Score: 1279.00 Matches: 245

Percent Similarity: 88.40% Conservative: 37

Best Local Similarity: 76.80% Mismatches: 37

Query Match: 79.00% Indels: 0

DB: 19 Gaps: 0

US-10-606-300-11 (1-319) x US-10-424-599-119463 (1-1361)

Qy 1 MethThrLysIleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGly 20
Db 49 CTTAGAGAAATGGAACTACGTGAGCTTGGAAAGAACCGGACTGAAACTCAGCACTGTAGGA 108
Qy 21 PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAla 40
Db 109 TTCGAGGCTCTCCAACTGGGCAATGCTTCGGCGATGTTCCGAGGAACAAGCCACGCT 168
Qy 41 ThrValArgGluAlaPheArgLeuGlyIleAsnPheAspThrSerProTyrTyrGly 60

Db 169 TCCGTTTCGCTCGCTTTTCCAAATCCGGCATCAATTTCTTCGACACTTCTCCGTACTACGGA 228
Qy 61 GlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSer 80
Db 229 GGGACACTGTTCGGAAGAAGTTCTGGGAAGGCACTGAAGGCTCTGGGAGCTCCGAGGAAC 288
Qy 81 AspTyrIleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGlu 100
Db 289 TCGTACGTTGTGGCAACGAGTGGCGACGGGTACAGGAGGGTTCGATTTCAGCGCGGAG 348
Qy 101 ArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 120
Db 349 AGAGTGACGAGAACATTTGAAGAGAGCTTGGAAAGGTTGCAGCTTGACTACGTCGACATT 408
Qy 121 LeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePro 140
Db 409 CTCCAAATGCCACGACATCGAGTTCGGCTCTTACAGCAGGTTGTGAATGAGACGATTCCA 468
Qy 141 AlaLeuGlnLysLeuLysGlnGluGlyThrArgPheIleGlyIleThrGlyLeuPro 160
Db 469 GCGCTTGTGAAGCTGAAGGAGCAGGGAAGCGCGGTTTCATTGGAATAACAGGGCTTCT 528
Qy 161 LeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeu 180
Db 529 CTGGGAGATTTTCAGTTACGTCTTGATAGGGTTCCGCTGGGACGCTGGATGTTGTGCTT 588
Qy 181 SerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLys 200
Db 589 TCATATTTGCCATTAATCTGTGTGAATGACACTAGCTTGGGGAATTTGGTCCCTATTTGAAG 648
Qy 201 SerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGln 220
Db 649 ACCAAAGGGTTGGCATTATCAATGCTTCTCCCTTGTCTATGGGCTTCTCACCGAGTCT 708
Qy 221 GlyProProGluTyrHisProAlaSerProGluLeuLysSerAlaSerLysAlaVal 240
Db 709 GGGCGGCTGAATGGCATCCGCACTCACTGGAATCTCAAGTCTGCAATGTCAGCTGCTCT 768
Qy 241 AlaHisCysLysSerLysGlyLysIleThrLysLeuAlaLeuGlnTyrSerLeuAla 260
Db 769 ACCATTTGTAAGAAAGAAAGAAACATTTCAAGTTAGATTTGATGATGATGATTTGTTA 828
Qy 261 AsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGln 280
Db 829 AATAAGGAATACATCAGTCTTGTAGGCAATGAAGTCTGTTGAACAGGTGGAGGAAAT 888
Qy 281 ValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnThrLeuSerGluVal 300
Db 889 GTTCTGCGCAGAGAACTTGCAACTTCTGGAATTTGATGAAGAGACTCTCTCTGAAGTC 948
Qy 301 GluAlaIleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGlnAsn 319
Db 949 GAAGCTATTCGAGCCTGTTTAAACACAGTCGTGGCCAGTGGATCCAGCAGAGC 1005

RESULT 4

US-10-437-963-11457

; Sequence 11457, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

```
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 11457
; LENGTH: 1616
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_17680C.1
US-10-437-963-11457

Alignment Scores:
Pred. No.:      2,34e-126      Length:      1616
Score:          1213.00      Matches:      234
Percent Similarity: 83.48%      Conservative: 29
Best Local Similarity: 74.29%      Mismatches: 52
Query Match:      20          Indels:      0
DB:              20          Gaps:      0

US-10-606-300-11 (1-319) x US-10-437-963-11457 (1-1616)

QY 4 IleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGlyPheGlyAla 23
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
348 ATGGAGCTCCGCGAGCTCGCGCCACAGCGGCTCCGCGTCAGCGCGCTCGGCTTCGGCGCC 407
QY 24 SerProLeuGlySerValPheGlyProValAlaGluAspAlaValAlaThrValArg 43
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
408 TCCCTCTCGGCGCATGTCTTCGGCGACGTCCCGCGACGTCCCGCGCGCGCGCGCTCCGC 467
QY 44 GluAlaPheArgLeuGlyLeuAsnPheAspThrSerProTyrTyrGlyGlyThrLeu 63
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
468 CGCGCCTCGACCTCGGCATCACTTCTTCGACACTCCCGCTACTACGGCGGCACGGTG 527
QY 64 SerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAspTyrIle 83
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
528 TCGGAGTCGGTCTCGGCGACTGCTCCGCGCGCGCGCGCTCCCGCGGACCGGTTCGTG 587
QY 84 ValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGluArgValArg 103
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
588 GTCCGCACCAAGTGGCGCGCTACAGGAAGGTTCGACTTCAGCGCGCGCGCGCTACCC 647
QY 104 LysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIleLeuHisCys 123
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
648 CGCAGCTCGAGGAGCGCTCGCAGCGCTCGGCTCGACTAGTCGACATCTCCACTGC 707
QY 124 HisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIleProAlaLeuGln 143
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
708 CACGACATCGAGTTCACCGACCTCGACAGATGTGAATGAGACGATTCGGTGTCTCCAG 767
QY 144 LysLeuLysGlnGluGlyThrArgPheIleGlyIleThrGlyLeuProLeuAspIle 163
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
768 AAGATCAAGGAGCGGGAAGCGGTTTCATCGGGATAACCGGCTCGCATTTGAGCATC 827
QY 164 PheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeuSerTyrCys 183
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
828 TACACTTATGTCTGACCAAGTGGCGCAGGTCTGTGGATGTGATCTGTCTTACTGC 887
QY 184 HisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrTyrLeuSerLysGly 203
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
888 CACTATGGGATCAATGATACCGCGTTGTGGATTGTCTTCCCTACATGAAGACCAAGGT 947
QY 204 ValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGlnGlyProPro 223
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
948 GTTGGCGTGATCAGTGCTTCACCTCTTGCAATGGGGCTTCCTTACAGATAACGGGCACCC 1007
QY 224 GluTrpHisProAlaSerProGluLeuLysSerAlaSerLysAlaAlaValAlaHisCys 243
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1008 GAATGGCACCTTCGCCGAAAGAACTAAAGTTGGCATGCGAGGCGAGCAGATCACTGT 1067
QY 244 LysSerLysGlyLysIleThrLysLeuAlaLeuGlnTyrSerLeuAlaAsnLysGlu 263
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1068 AAGAAGAGAGGGGAAAAACATTACAAAGCTAGCTATGTCAGTATGATGTAACCAATGAG 1127
QY 264 IleSerSerValLeuValGlyMetSerSerValSerGlnValGluGluAsnValAlaAla 283
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Db 514 CAGAAAGTCAAGAGGAATGGGAAGCACGGTTTCATTGGGCATCACCGGGTTGCCCTCTCAGC 573
Qy 163 IlePheThrTyrValLeuAspArgValProGlyThrValAspValIleLeuSerTyr 182
Db 574 ATCTACCCCTTATGCTCGACCGTGTACACACCGCTCGGTGGAGTGAATCTATCTAT 633
Qy 183 CysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuProTyrLeuLeuSerLys 202
Db 634 TGTCACTACGGGATCAATAGACACCTCCCTTGTGATCTGCTCCCTACTTGAAGAGCAAA 693
Qy 203 GlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGlnGlyPro 222
Db 694 GGTGTGGGGTATCAGTGGCTTCGGCCCTCTCCATGGGTCTTTTAACAGATAAGGGCCA 753
Qy 223 ProGluTyrHisProAlaSerProGluLeuLysSerAlaSerLysAlaAlaValAlaHis 242
Db 754 CCGGAGTGGCACCTCGCACAGAGAACTTAAGTCAGCATGCNAGGCTGCAGCAGATCAC 813
Qy 243 CysLysSerLysGlyLysIleThrLysLeuAlaLeuGlnTyrSerLeuAlaAsnLys 262
Db 814 TGTAGAAAGAGGGGAAAAGCAATACGAAGCTAGCTATGCAGTACAGCTTAATGAACAAT 873
Qy 263 GluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluAsnValAla 282
Db 874 GAAATCTCGACAGTTCTTGTGGATGAACCTTTTGGAAACAGGTGGAGGAGAAATGTGGCG 933
Qy 283 AlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluValGluAla 302
Db 934 GCTGCACCTGGAGTTCTCAATCAGCGCATCGATCGATGANGAACTTATGCGTGAAGTTGAAGCA 993
Qy 303 IleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGln 318
Db 994 ATTCCTGAGCCTGTGAAGAATCTGACATGGCTAGCGGCATCCCAACAA 1041

RESULT 6

US-10-767-701-13219
; Sequence 13219, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 13219
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS9142_1
US-10-767-701-13219

Alignment Scores:
Pred. No.: 6,59e-125 Length: 1434
Score: 1199.50 Matches: 233
Percent Similarity: 83.86% Conservative: 32
Best Local Similarity: 73.73% Mismatches: 50
Query Match: 74.03% Indels: 1
DB: 20 Gaps: 1

US-10-606-300-11 (1-319) x US-10-767-701-13219 (1-1434)

Qy 4 IleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGlyPheGlyAla 23
Db 55 ATGGAGCTCCGGAGCTGGCGGTACGGGCTCCAGTCAGCGCTGGCTTTGGCGCC 114
Qy 24 SerProLeuGlySerValPheGlyProValAlaGluAspAlaValAlaThrValArg 43
Db 115 TCCCCGCTCGGCAACGTTTTCGGGACGTCGCCCGCATACCGCCCGCGCGCGCTCGCG 174

Qy 44 GluAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyrTyrGlyGlyThrIleu 63
Db 175 CGCGCGCTCGACCTCGGCATCAACTTCTTCGACACCTCCCGTACTACGGGGCAGCGTC 234
Qy 64 SerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAspTyrIle 83
Db 235 TCGGAGTCAGTCCCTCGCGGATTCCTCGCCACCGCGCGCTTCGCGGAGCCGAGTCGTC 294
Qy 84 ValAlaThrLysCysGlyArgTyrLys---GluGlyPheAspPheSerAlaGluArgVal 102
Db 295 GTCCGCCACCAAGTCGGCGCGCTACAAAGACGAGGGTTTCGACTTCTCCCGCGACCGTGTG 354
Qy 103 ArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIleLeuHis 122
Db 355 ACGCGCAGCATAGACAGAGCTCGCGCGCTGGGGCTGGGACTACGTTGACATCTCTCCAC 414
Qy 123 CysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIleProAlaLeu 142
Db 415 GCCCAGCAGATCGAGTTTCAACCATCTCGACAGATTTGGAATGAGACAAATTCGCCGCTC 474
Qy 143 GlnLysLeuLysGlnGluGlyThrArgPheIleGlyIleThrGlyLeuProLeuAsp 162
Db 475 CAGAAGATTAGGAGAGTGGGAGGACGCTTTCATTGGCATCACCGGGCTGCCCTCAGC 534
Qy 163 IlePheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeuSerTyr 182
Db 535 ATCTACCCCTTATGCTCTCGACAGGTAGCACACCGCTCGGTGGACCTGATCTATCTTAC 594
Qy 183 CysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLysSerLys 202
Db 595 TGCACACTACGGGATCAATGACACCTCCCTTGTGATCTGCTCCCTACTTGAAGAGCAAA 654
Qy 203 GlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGlnGlyPro 222
Db 655 GGTGTGGGGTATCAGTGGCTTCGCCACTCTCCATGGGTCTTTTAAACAGATAATGGCCA 714
Qy 223 ProGluTyrHisProAlaSerProGluLeuLysSerAlaSerLysAlaAlaValAlaHis 242
Db 715 CCGGAGTGGCACCTCGCACAGAACTTAAGTCAGCATCGAGGCTGCAGCAGATCAC 774
Qy 243 CysLysSerLysGlyLysIleThrLysLeuAlaLeuGlnTyrSerLeuAlaAsnLys 262
Db 775 TGTAGAAAGAGGGGAGAGCAATACCAAGCTAGCTATGCAGTATAGCTTGAATGAATAT 834
Qy 263 GluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGluAsnValAla 282
Db 835 GAAATCTCGACATTTCTTGTGGATGAATGAACTCTTCAGAACAGGTGGAGGAGAAATGTGGCT 894
Qy 283 AlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluValGluAla 302
Db 895 GCTGCACCTGGAGTTGTCAACATCAGGCATTCATGAAGAACTTCTCGGTGAAGTTGAAGCA 954
Qy 303 IleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGln 318
Db 955 ATTCCTGAGCCTGTGAAGAACCTGACTTGGCTAGCGGCATCCCAACAA 1002

RESULT 7

US-10-425-115-95234
; Sequence 95234, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 95234

Qy	253	LeuAlaLeuGlnTyrSerLeuAlaAsnLysGluIleSerSerValLeuValGlyMetSer	272
Db	781	CTAGCTATGCAGTACAGCTTTAATGAACAATCGAACATCTCGACAGTGCTTGTGGTAATGAAC	840
Qy	273	SerValSerGlnValGluGluAsnValAlaValThrGluLeuGluSerLeuGlyMet	292
Db	841	TCVITGGAAACAGGTGGAGGAAATGTGGCGCTGCCACTGGAGTTGTCAACATCAGGCATC	900
Qy	293	AspGlnGlnThrLeuSerGluValGlulAlaIleLeuGluProValLysAsnLeuThrTrp	312
Db	901	GATCATGAACCTATTATGCGTGAAGTTGAAGAATACTCTGAGCGCTGTGAAGAATCTGCATGG	960
Qy	313	ProSerGlyIleHisGln	318
Db	961	CCTAGCGGCATCCAACAA	978
RESULT 8			
US-10-425-114-31543			
; Sequence 31543, Application US/10425114			
; Publication No. US2004003488A1			
GENERAL INFORMATION:			
; APPLICANT: Liu, Jingdong			
; APPLICANT: Zhou, Yihua			
; APPLICANT: Kovacic, David K.			
; APPLICANT: Screen, Steven E			
; APPLICANT: Tabaska, Jack E			
; APPLICANT: Cao, Yongwei			
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With			
; FILE REFERENCE: 38-21(5331)B			
; CURRENT APPLICATION NUMBER: US/10/425,114			
; CURRENT FILING DATE: 2003-04-28			
; NUMBER OF SEQ ID NOS: 73128			
; SEQ ID NO 31543			
; LENGTH: 1134			
; TYPE: DNA			
; ORGANISM: Zea mays			
; FEATURE:			
; OTHER INFORMATION: Clone ID: UC-ZMFLB73186H07_FLI			
US-10-425-114-31543			
Alignment Scores:			
Pred. No.:	4,29e-92	Length:	1134
Score:	908.00	Matches:	174
Percent Similarity:	86.73%	Conservative:	22
Best Local Similarity:	76.99%	Mismatches:	30
Query Match:	56.08%	Indels:	0
DB:	19	Gaps:	0
US-10-606-300-11 (1-319) x US-10-425-114-31543 (1-1134)			
Qy	93	GluGlyPheAspPheSerAlaGluAtqValArgLysSerIleAspGluSerLeuGluArg	112
Db	9	GAGGGTTTCACATTCTCCGCCAACCGGTGACACGACGATAGACGAGAGCTCGCCCGA	68
Qy	113	LeuGlnLeuAspTyrValAspIleLeuHisCysHisAspIleGluPheGlySerLeuAsp	132
Db	69	CTGGGGCTGGACTACGTTCGACATCTCCACGCTCATGACATCGAGTTCACCATCTCGAA	128
Qy	133	GlnIleValSerGluThrIleProAlaLeuGlnLysLeuLysGlnGlyLysThrArg	152
Db	129	CAGATTGTGAATGAGACAAATCCCGCACCTCCAGAAGATCAAGGAGAATGGGAAGCACGG	188
Qy	153	PheIleGlyIleThrGlyLeuProLeuAspIlePheThrTyrValLeuAspArgValPro	172
Db	189	TTCATTGGCATCACCGGGTTCCTCTCAGCATCTTACCCCTTATGCTCGACCGGTGAGCA	248
Qy	173	ProGlyThrValAspValIleLeuSerTyrCysHisTyrGlyValAsnAspSerThrLeu	192
Db	249	CCAGGCTCGGTGACAGTGATTCTATCTTATGCTACTACGGGATCAATGACACCTCCCTT	308
Qy	193	LeuAspLeuLeuProTyrLeuLysSerLysGlyValGlyValIleSerAlaSerProLeu	212

```
Db 309 GTTGATCTGCTCCCTACTTGAAGAGCAAGGTGTGGGGTTATCAGTGTGGCCCTC 368
Qy 213 AlaMetGlyLeuLeuThrGluGlnGlyProGluTyrHisProAlaSerProGluLeu 232
Db 369 TCCATGGGTCTTTTAAACAGATAATGGCCACCGGAGTGGCCACCTGCGACCAAGAACTT 428
Qy 233 LysSerAlaSerLysAlaAlaValalaHisCysLysSerLysGlyLysLysLysLys 252
Db 429 AAGTCAGCATGCAAGCTGCGACGACATCACTGTAGAAAAGAGGGGAAAGCAATTACGAAG 488
Qy 253 LeuAlaLeuGlnTyrSerLeuAlaLeuLysGlyLysLysSerValLeuGlyMetSer 272
Db 489 CTAGCTATGCGATACAGCTTATGACCAATGAATCTCGACAGTCTCTTGTGGAAATGAC 548
Qy 273 SerValSerGlnValGluGlnValAlaAlaValThrGluLeuGluSerLeuGlyMet 292
Db 549 TCTTTGGAACAGGTGGAGGAGAAATGTGGCGCTGCACTGGAGTTGTCAACATCAGGCATC 608
Qy 293 AspGlnGluThrLeuSerGluValGluAlaLeuGluProValLysAsnLeuThrTyr 312
Db 609 GATGAAGAACTTATCGTGAAGTTGAAGCAATTCGAGCCCTGTGAAGAATCTGCACATGG 668
Qy 313 ProSerGlyIleHisGln 318
Db 669 CCTAGCGGCATCCAAACA 686

RESULT 9
US-10-021-323-3555
; Sequence 3555, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US/10/021,323
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255, 619
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 3555
; LENGTH: 564
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(564)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3825-027-Q6-K6-G2
US-10-021-323-3555

Alignment Scores:
Pred. No.: 519e-78 Length: 564
Score: 779.50 Matches: 155
Percent Similarity: 88.8% Conservative: 12
Best Local Similarity: 82.45% Mismatches: 20
Query Match: 48.15% Indels: 2
DB: 20 Gaps: 1

US-10-606-300-11 (1-319) x US-10-021-323-3555 (1-564)

Qy 66 LysMetLeuGlyLysGlyLysAlaLeuGlnValProArgSerAspTyrIleValala 85
Db 3 CGANAGCTGGTAAAGGACTTAAAGCTCTTGGAGTCTCTAGAAATGAATATATGTTTC--- 59
Qy 86 ThrLysCysGlyArgTyrLysGlyLysPheAspPheSerAlaGluArgLysSer 105
Db 60 GCAAAATGTGGAGATATCGTGAAGGTTTGTATTTTCTAGTCTGAGAGATTAACATAAGC 119
Qy 106 IleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAlaPheIleLeuHisCysHisAsp 125
```

```
Db 120 ATTGATGAAGAGCTGGAGAGGTTCACACTTGATTTGATATATTTCAATGCCATGAC 179
Qy 126 IleGluPheGlySerLeuAspGlnIleValSerGluThrIleProAlaLeuGlnLysLeu 145
Db 180 ATTGAATTCGGCTCTCTTGATCAGTGTGTGAATGAGAGATTCGGCACTTCANAACTG 239
Qy 146 LysGlnGluGlyLysThrArgPheIleGlyLysLeuThrGlyLeuProLeuAspIlePheThr 165
Db 240 AAGCAAGCAGGAAGATTCGTTTCATTGGTATCACCGGGTTCGCCCTTGGAAATTTTACT 299
Qy 166 TyrValLeuAspArgValProGlyThrValAspValIleLeuSerTyrCysHisTyr 185
Db 300 TATGTGCTTGATAGGTTCACCGAGGCACTGTGATGTAATATATATATATGTCATTAT 359
Qy 186 GlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLysSerLysGlyValGly 205
Db 360 AGCATTATGATTCACATCTGGAGATTTATGCTTACTTGTAAACCAAGT-GTTGGC 418
Qy 206 ValLeuSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGlnGlyProGluTyr 225
Db 419 GTAAATCAGTCATCTCCACTTGTATGGCACTTCTTACTGAGTTGGTCCACCGAGTGG 478
Qy 226 HisProAlaSerProGluLeuLysSerAlaSerLysAlaAlaValAlaHisCysLysSer 245
Db 479 CATCGGCATCTCCGAACTCAAGTCTGCCTGCCAAGCTGCTGTATATTTGTAAGAAG 538
Qy 246 LysGlyLysLysIleThrLysLeu 253
Db 539 AAAGAAAGAAATATTTGCAAGTTA 562

RESULT 10
US-11-097-143-28022
; Sequence 28022, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28022
; LENGTH: 1038
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-28022

Alignment Scores:
Pred. No.: 1.38e-56 Length: 1038
Score: 593.50 Matches: 128
Percent Similarity: 63.46% Conservative: 63
Best Local Similarity: 42.52% Mismatches: 99
```

```
Query Match: 36.66% Indels: 11
DB: 26 Gaps: 7
US-10-606-300-11 (1-319) x US-11-097-143-28022 (1-1038)
QY 3 LysIleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGlyPheGly 22
DB 61 CGGATGGAATATCGCAATCTCGGAAGACCGCGCTCGCAAGTCTCGAAAGTCTCTTTCCGA 120
QY 23 AlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAlaThrVal 42
DB 121 GCGGCGGCTGTGCGGACGACGCTGTTGATTTGGAG--GAGGTATTAACACTGTG 177
QY 43 ArgGluAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyrTyrGlyGlyThr 62
DB 178 CACGAGGCGGTAAGTCAGGCACTCAACTACATTGACACTGCTCCCTGGTATGGTCAGGT 237
QY 63 LeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAspTyr 82
DB 238 CGCTCTGAGGAGTCTGCGGACTGCGCCCTAAAG-----GATGTGCCGCGGAATCCTAC 291
QY 83 IleValAlaThrLysCysGlyArgTyrLys-----GluGlyPheAspPheSer 98
DB 292 TATATCGCCACGAAGTCTGCTCGTACGNACTGGACTAGATAAATGTTGACTTTAGT 351
QY 99 AlaGluArgValArgLysSerIleAspLeuGluArgLeuGlnLeuAspTyrVal 118
DB 352 GCCAAGACGCGCGAAGCGTGGAGAGAGCTTGAACTACTTTGGCTGGACTACGTT 411
QY 119 AspIleLeuHisCysHisAspIleGluPheGly---SerLeuAspGlnIleValSerGlu 137
DB 412 GATGTCATCCAGATTACAGATTCAGATTGCGCAAGGATCTGGACATTGTGATCAACGAG 471
QY 138 ThrIleProAlaLeuGlnLysGlyGlnGlyLysThrArgPheIleGlyIleThr 157
DB 472 ACATGCCACCTTGGAGCAGCTGTCAAGAGGGGCAAGGCAAGATTCAATGGATGCC 531
QY 158 GlyLeuProLeuAspIlePheThrTyrValLeuAspArgValProGlyThrValAsp 177
DB 532 GCTTACCGGATTCGCTGCTTAAGAGATTCTCTGACCCGGAACA---GCCGGAAGACTCGAT 588
QY 178 ValIleLeuSerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuPro 197
DB 589 ACGGTCTCCTACCTATGCCAGATACACCTGACCGATGAACGCTCTGTGAGTACCTGGAT 648
QY 198 TyrLeuLysSerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeu 217
DB 649 TTCTTCAGTCCAGAACCTGGCGTCTATCTGTGCGGAGCTCATGCTCGACTGCTG 708
QY 218 ThrGluGlnGlyProGluTrpHisProAlaSerProGluLeuLysSerAlaSerLys 237
DB 709 ACCAATGCCGCTCCAGCCATGCGATCGCGGCTGATGATGACAGAGGCCATTCGCCGG 768
QY 238 AlaAlaValAlaHisCysLysSerLysGlyLysLysIleThrLysLeuAlaLeuGlnTyr 257
DB 769 AAGGCATCGAGGTCTGCAAGAACCGCGCTGGAGCTGGCGCAAGCTGCTGCTACTACTAC 828
QY 258 SerLeuAlaAsn---LysGluIleSerSerValLeuValGlyMetSerSerValSerGln 276
DB 829 ACAGTGGGCGGACTGCCGAGTGGAGCCTTCCTAACCGGGATCGACGAGCGCCAGTTG 888
QY 277 ValGluGluAsnValAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThr 296
DB 889 CTGCGAATCAACTGTGATGCC---AACGAAGTGGGCTCAGCGCATAAAGGAGCAGGAAGTG 945
QY 297 Leu 297
DB 946 TTG 948
RESULT 11
US-11-097-143-28021/c
; Sequence 28021, Application US/11097143
; Publication No. US2005020858A1
```

```
GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; PRIOR FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28021
; LENGTH: 3724
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-28021
Alignment Scores:
Pred. No.: 1,896-50 Length: 3724
Score: 546.50 Matches: 131
Percent Similarity: 51.19% Conservative: 63
Best Local Similarity: 34.56% Mismatches: 97
Query Match: 33.76% Indels: 88
DB: 26 Gaps: 9
US-10-606-300-11 (1-319) x US-11-097-143-28021 (1-3724)
QY 3 LysIleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGlyPheGly 22
DB 2664 CGGATGGAATATCGCAATCTCGGAAGACCGCGCTCGCAAGTCTCGAAAGTCTCTTTCCGA 2605
QY 23 -AlaSerProLeuGlySerValPheGly----- 31
DB 2604 GCGCGCGCCCTGTGCGCGAATCTACGGGTAAGAGAAATCTTAACAAATTCATTTACTCCGG 2545
QY 31 ----- 31
DB 2544 TACACAGAAAGATATTTCTTCTTCTTGGCATTTAATAATAAATTTCTAGAAAGTATTG 2485
QY 31 ----- 31
DB 2484 CTGAAAGTACGAATTCGTAATTTAAATAATGCGGAAATGTTTGAGATTGTATTTCATCT 2425
QY 32 -----ProValAlaGluAsp-----AspAlaValAlaThrValArgG1 44
DB 2424 CTAATTCATATATTTCCCTTTTCAGTTTGTGAGGAGGGTATTAAACCTGTGACGA 2365
QY 44 uAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyrTyrGlyGlyThrLeuSe 64
DB 2364 GCGCGTAAGTCAGGCACTCACTACATTGACACTGCTCCCTGGTATGGTCAGGGTCGCTC 2305
QY 64 rGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAspTyrIleVa 84
DB 2304 TGAGGAGGTCTCTGGGACTGGCCCTAAAG-----GATGTGCCGCGGGAATCCTACTATAT 2251
QY 84 lAlaThrLysCysGlyArgTyrLys-----GluGlyPheAspPheSerAlaG1 100
```



```
Db 2250 GCCACGAAAGTGGCTCGCTACGAACCTGGACTACGATAAATGTTTGACTTTAGTGCCAA 2191
Qy 100 uArgValArgYsHisAspIleGluPheGly---SerLeuAspGlnLeuValAspIle 120
Db 2190 GAAGACGCGGAAAGCGTGAGAGAGCTTGAACACTTGGCTGGACTACGTTGATG 2131
Qy 120 eLeuHisCysHisAspIleGluPheGly---SerLeuAspGlnLeuValSerGluThrIle 139
Db 2130 CATCCAGATTACGATATCGAGTTCGCCAAGGATCTGGACATTTGTGATCAACGAGACACT 2071
Qy 139 eProAlaLeuGlnLeuLysGlnGluGlyThrArgPheIleGlyIleThrGlyLe 159
Db 2070 GCCCACTTTGGAGCGAGCTGGTCAGAGGCGGAGCAAGATTCATGAGGTGTCGGCTTA 2011
Qy 159 uProLeuAspIlePheThrTyrValLeuAspArgValProGlyThrValAspVal-- 178
Db 2010 CCGGATTTCCGGTCTTAAGGAGTTCCTGCACCGAACA---GCCGGAAGACTCGATGTGAG 1954
Qy 179 -----IleLeu 180
Db 1953 TTTATAGCATACACTTCAAAATTTCAAGTTTAAATCTCATATCGCCACCAGACGGTCTCT 1894
Qy 180 uSerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeu 200
Db 1893 CACCTATGCCAGATACACCCGACGATGAACGCTCCTGGAGTACCTGGATTCTTCAA 1834
Qy 200 sSerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluG1 220
Db 1833 GTCCAGAACCTGGCGGTCACTGTGGCGCAGCTCATGCCCTCGACTGCTGACCAATGC 1774
Qy 220 nGlyProGluThrHisProAlaSerProGluLeuLysSerAlaSerLysAlaAlaVa 240
Db 1773 CGGTCCACAGCCATGCCATCCGGCCAGTGCATGATGAGCAGAGGCCATTCGCCGAGGCCATC 1714
Qy 240 lAlaHisCysLysSerLysGlyLysLysIleThrLysLeuAlaLeuGlnTyrSerLeuAl 260
Db 1713 GGAGCTGCAAGGAAGACCGCGGTGGAGCTGGGCAAGCTGGCCATGTACTACACATGAG 1654
Qy 260 aAsn---LysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluG1 279
Db 1653 CGGACTGCGGAGTGGAGCACCTTCTTAACGGGCATGACAGCGCCAGTGTGTCGGAT 1594
Qy 279 uAsnValAlaAlaValThrGluLeuSerLeuGlyMetAspGlnGluThrLeu 297
Db 1593 CAACCTGATGCGC---AACGAAGTGGCGCTCAGCGATGAAGGAGCAGGAAGTGTG 1542

RESULT 12
US-11-097-143-40478
; Sequence 40478, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
```

```
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40478
; LENGTH: 1051
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-40478

Alignment Scores:
Pred. No.: 4,98-51 Length: 1051
Score: 544.50 Matches: 117
Percent Similarity: 58.75% Conservat: 71
Best Local Similarity: 36.56% Mismatches: 117
Query Match: 33.63% Indels: 15
DB: 26 Gaps: 7

US-10-606-300-11 (1-319) x US-11-097-143-40478 (1-1051)
```

```
Qy 3 LysIleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGlyPheGly 22
Db 83 CGTATGGAGTACCGCCAACTCGGATCCACGGGTCTGCGAGTATCCAAAATGCTTTGGGT 142
Qy 23 AlaSerProLeuGlySerValPheGly---ProValAlaGluAspAspAlaValAlaThr 41
Db 143 GGTGCCACTCTCTCCAAAGCTCTTCCGATGACTTCGATCGCAGAGGGGATCTCTCAGC 202
Qy 42 ValArgGluAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyrTyrGlyGly 61
Db 203 GTGCAGGAAGCCATTAGATCCGGTATCAACTACATAGACACGGCTCCCTTTATGGCCAA 262
Qy 62 ThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAsp 81
Db 263 GGCAAATCGGAAGAGCTGCTTGGCCAGCGCTCAAG-----GATGTGCCCGGAGGCC 316
Qy 82 TyrIleValAlaThrLysCysGlyArgTyrLys-----GluGlyPheAspPhe 97
Db 317 TATTATATAGCAACTAAAGATTGCACGTTACGAGTTGGATGGATCCAAACAATATGTTCGACT 376
Qy 98 SerAlaGluArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyr 117
Db 377 ACGGCTGCCAAGGCTCGGGAGAGTGTGAAGCGTGTCTGGAGCTGCTCCAGTTGGACAGG 436
Qy 118 ValAspIleLeuHisCysHisAspIleGluPheGly---SerLeuAspGlnIleValSer 136
Db 437 GTGACGCTACTACAGTTTCATGACGTGATCGCGCACCTAGTCTGGACATGCTGTGAT 496
Qy 137 GluThrIleProAlaLeuGlnLysGlnGluGlyLysThrArgPheIleGlyIle 156
Db 497 GAGACCATACCGCTCTCGAGAGTACGTCAGGCGGGAAGGCTCGATTCATCGAGGTC 556
Qy 157 ThrGlyLeuProLeuAspIlePheThrTyrValLeuAspArgValProGlyThrVal 176
Db 557 ACCGCTACGATGTGAGACGTCTGAAGGAGTGTGCGGAGCGG---GGCAAAGGTCCGATC 613
Qy 177 AspValIleLeuSerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeu 196
Db 614 CAGTGTGTCTCAACTATGCCCGCTACACCTGTGTGGACAACACTTTGCTGGCCACATG 673
Qy 197 ProTyrLeuLysSerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeu 216
Db 674 AAGGCTTCCAGGAGATGGGCGTGTGCGGCTTCTGTGCGGCGCGCACTCTCTGGGATC 733
Qy 217 LeuThrGluGlnGlyProGluThrHisProAlaSerProGluLeuLysSerAlaSer 236
Db 734 CTAAAGCAACGCTGGACCCCACTCTGCTGTGATGTGCGGAACCTCTCTAGCTGTGGGC 793
Qy 237 LysAlaAlaValAlaHisCysLysSerLysGlyLysIleThrLysLeuAlaLeuGln 256
Db 794 AAACGGGAGCGGAAATCTGTGAGAGGAGCAAGCTGTGAGCTTGGAAAGCTGGCCATG 853
Qy 257 TyrSerLeuAlaAsnLysGluIleSerSerValLeuValGlyMetSerValSerGln 276
```



```
/ ORGANISM: Gossypium hirsutum
/ FEATURE:
/ OTHER INFORMATION: Clone ID: LIB3825-027-Q6-N6-G2
US-10-021-323-3629

Alignment Scores:
Pred. No.: 4,586-46 Length: 563
Score: 497.00 Matches: 97
Percent Similarity: 88.00% Conservatives: 13
Best Local Similarity: 77.60% Mismatches: 15
Query Match: 30.70% Indels: 0
DB: 20 Gaps: 0

US-10-606-300-11 (1-319) x US-10-021-323-3629 (1-563)
QY 195 LeuLeuProTyrLeuLysSerLysGlyValGlyValIleSerAlaSerProLeuAlaMet 214
DB 561 TTAATGCCCTACTTGAACCAACCAAGGTGTGGCGTAATCAGTCATCCATCTGCTATG 502
QY 215 GlyLeuLeuThrGluGlnGlyProProGluTTPHisProAlaSerProGluLeuLysSer 234
DB 501 GGACTTCTTACTAGTTGGTCCACCGGAGTGGCATCCGGCATCTCCGCACTCAAGTCT 442
QY 235 AlaSerLysAlaAlaValAlaHisCysLysSerLysGlyLysLysLysLysLysLys 254
DB 441 GCCTGCCAAGCTGCTGTATATTTGTAAGAGAGAAAGAAAGAAATATTTTCAAGTTAGCT 382
QY 255 LeuGlnTyrSerLeuAlaAsnLysGluIleSerSerValLeuValGlyMetSerSerVal 274
DB 381 ATGCATACAGCTTGTCAACCAAGATATTTTCGACAGTGGCTGGTGGCATGAATCTCGGTT 322
QY 275 SerGlnValGluGlnValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGln 294
DB 321 AAACAGGTTGAAGAGAAATGTTGCCGCTGCAACAGAACTTGCACATTCGGGAAAGATCAT 262
QY 295 GluThrLeuSerGluValGluAlaIleLeuGluProValLysAsnLeuThrTTPProSer 314
DB 261 GAAACTCTAGCTAGGTGGAGCAATACATACTGAAGCCAGTGAAGCAATCAGACATGCGCGAGC 202
QY 315 GlyIleHisGlnAsn 319
DB 201 GGAATTCACGAAGC 187

RESULT 15
US-11-097-143-7144/c
/ Sequence 7144, Application US/11097143
/ Publication No. US20050208558A1
/ GENERAL INFORMATION:
/ APPLICANT: Venter, J. Craig
/ TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
/ TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
/ TITLE OF INVENTION: DROSOPHILA GENES.
/ FILE REFERENCE: CL000728
/ CURRENT APPLICATION NUMBER: US/11/097,143
/ PRIOR FILING DATE: 2005-04-04
/ PRIOR APPLICATION NUMBER: 60/157,832
/ PRIOR FILING DATE: 1999-10-05
/ PRIOR APPLICATION NUMBER: 60/160,191
/ PRIOR FILING DATE: 1999-10-19
/ PRIOR APPLICATION NUMBER: 60/161,932
/ PRIOR FILING DATE: 1999-10-28
/ PRIOR APPLICATION NUMBER: 60/164,769
/ PRIOR FILING DATE: 1999-11-12
/ PRIOR APPLICATION NUMBER: 60/173,383
/ PRIOR FILING DATE: 1999-12-28
/ PRIOR APPLICATION NUMBER: 60/175,693
/ PRIOR FILING DATE: 2000-01-12
/ PRIOR APPLICATION NUMBER: 60/184,831
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: 60/191,637
/ PRIOR FILING DATE: 2000-03-23
/ NUMBER OF SEQ ID NOS: 43008

/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 7144
/ LENGTH: 3014
/ TYPE: DNA
/ ORGANISM: DROSOPHILA
US-11-097-143-7144

Alignment Scores:
Pred. No.: 1,41e-40 Length: 3014
Score: 458.00 Matches: 116
Percent Similarity: 54.30% Conservatives: 67
Best Local Similarity: 34.42% Mismatches: 115
Query Match: 28.29% Indels: 41
DB: 26 Gaps: 8

US-10-606-300-11 (1-319) x US-11-097-143-7144 (1-3014)
QY 3 LysIleGluLeuAraGluAlaLeuGlyAsnThrGlyLysValSerAlaValGlyPheGly 22
DB 1954 CGTATGGAGTACCGCAACTCGGATCCACGGTCTGCACGTCTCTAAATTAGCTTAGGT 1895
QY 23 AlaSerProLeuGlySerVal---PheGlyProValAlaGluAspAlaValAlaThr 41
DB 1894 GGCTCCCCCTTTCGAATCTCTTCTTGATGACTATGATCGGAGGGGCACTCTTATG 1835
QY 42 ValArgGluAlaPheArgLeuGlyIleAsnPheAspThrSerProTyrTyrGlyGly 61
DB 1834 GTGCAGGAGGCCATTAGATCTGGCATTAACTATATATAGACACAGCTCCATCTTATCG--- 1779
QY 62 ThrLeuSerGluLysMetLeuGlyLysGlyLysAlaLeuGlnValProArgSerAsp 81
DB 1778 -----GAGGTGCTGCTTGGCCAGGCCCTAAAG-----GATGTACCCCGGAGGCC 1734
QY 82 TyrIleValAlaThrLysCysGlyArgTyr-----LysGluGlyPheAspPhe 97
DB 1733 TATTATATTGCACTAAAGTGGCGGCTACGGGCTGGATCCGAGAAATATGTTGACTAT 1674
QY 98 SerAlaGluArgValArgLysSerIleAspGluSerLeuGluGlnLeuAspTyr 117
DB 1673 TCGGCTGACAAAGCTCGGGAGAGTGTGAAGCGGAGTCTGGAGCGCTCCAGTTGGACAGG 1614
QY 118 ValAspIleLeu-----121
DB 1613 GTGGACATACATACAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1554
QY 122 -----HisCysHisAspIleGluPheGly---SerLeuAspGlnIleValSerGluTh 138
DB 1553 CTTATGACAGGTTTATGACGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1494
QY 138 rIleProAlaLeuGlnLysLeuLysGlnGluGlyThrArgPheIleGlyIleThrGl 158
DB 1493 CATACCGCTCTCGAGGAGTACGCTCCAGCGGGAAGGCTCGATTCACTCGGAGTCAACGC 1434
QY 158 yLeuProLeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAspVa 178
DB 1433 CTACGATGTGACAGCTGCTGAAGAGAGTGTGCGGAGCGG---GGCAAGGCTCGCATTCAGT 1377
QY 178 lIleLeuSerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuProTy 198
DB 1376 GGTCTCAACTATGCTGCTTACACCTTTTGAACACACCTTGTCTGCGCTACATCAAGGA 1317
QY 198 rLeuLysSerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuTh 218
DB 1316 CTTCCAGAAATATGGAGTGGGCGTGTCTGTGCGGCGGCTCACTCATTTGGGACTCTTGAG 1257
QY 218 rGluGlnGlyProProGluTTPHisProAlaSerProGluLeuLysSerAlaSerLysAl 238
DB 1256 AAACGCTGGACCAATGATGATCGCATCCCGGTAGTCAGGAAATCTCTGGCGCTGGCCAAACG 1197
QY 238 aAlaValAlaHisCysLysSerLysGlyLysLysLysLysLysLysLysLysLys 258
DB 1196 GGGGCGCGAAATCTGCCAGACAGAGCACTGGAGCTGGGAAAGCTGGCCATCTACTATAC 1137
```

```

QY 258 rLeuAlaAsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValG1 278
Db 1136 GATGCAACTGGATGGGGCGGCACCTTCCTCATCGGCATCCCAACCGAAAGCTGCTGCG 1077
QY 278 uGluAsnValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSe 298
Db 1076 GATTAACTGGACGCGATC-----TTCGACGGTCTCACTTCCCA 1038
QY 298 rGluValGluAlaIleLeuGluProValIysAsnLeuThrTrpProSer 314
Db 1037 CGAACAGGAAGTGTGCAGTATT-GCGCGAAAAGTAAGTTGGCCAACA 990

```

Search completed: November 13, 2005, 11:33:41
Job time : 917 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 13, 2005, 08:38:59 ; Search time 3686 Seconds
(without alignments)
3294.222 Million cell updates/sec

Title: US-10-606-300-11

Perfect score: 1619

Sequence: 1 MTKIEALRGNTGLKVSAGV.....VEAILBPVKLTPSGIHQN 319

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool_h/US10606300/runat_07112005_094938_8190/app.query.fasta_1.455
-DB=EST -QFMT=fastacp -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPT=0 -LOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=Blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US10606300 @CGN 1 1 3437 @runat_07112005_094938_8190 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DBLOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_ges1:*
9: gb_ges2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1367	84.4	894	5 BX839165	BX839165 BX839165
2	1308	80.8	953	7 CK253464	CK253464 EST737101
3	1262	77.9	998	7 CK251403	CK251403 EST735040
4	1246	77.0	987	7 CK250147	CK250147 EST733784
5	1232	76.1	948	7 CK243867	CK243867 EST727504
6	1212	74.9	979	7 CK282536	CK282536 EST745258
7	1208.5	74.6	1287	3 AY106400	AY106400 Zea mays
8	1207	74.6	1302	9 CL979301	CL979301 OSIFCC033
9	1198	74.0	945	7 CK245431	CK245431 EST729068

10	1189	73.4	936	7	CK257246	CK257246
11	1160	71.6	883	7	CK286730	EST749452
12	1147	70.8	922	7	CK284486	CK284486
13	1120	69.2	896	7	CK286441	EST749163
14	1092	67.4	926	7	CK282535	CK282535
15	1074	66.3	859	7	CK297635	CK297635
16	1069.5	66.1	822	7	CO366839	RTKL 350A
17	1069	66.0	803	7	CO121316	GR_EB02H
18	1061	65.5	828	6	CD576576	CD576576 UCRPT01.0
19	1053	65.0	638	1	AV825911	AV825911
20	1036	64.0	912	7	CK243866	CK243866
21	1032	63.7	814	7	CK294400	EST727503
22	1031	63.7	926	7	CK250148	EST757114
23	1018	62.9	810	6	CB618586	CK250148
24	1017	62.8	777	7	CK254163	CK254163
25	1015	62.7	826	7	CK297748	EST770462
26	1014	62.6	817	7	CK284947	CK284947
27	1012	62.5	695	7	CV430394	CK284947
28	1010	62.4	741	7	CF568880	CK284947
29	1005	62.1	706	4	BG646660	CK284947
30	997.5	61.6	804	7	CN130033	CK284947
31	995	61.5	698	4	B1920114	EST540049
32	991	61.2	719	7	CF417257	USDA-FP.1
33	988.5	61.1	834	7	CO363121	CK284947
34	971	60.0	670	5	B0888949	CK284947
35	960	59.3	858	7	CK282537	EST745259
36	950	58.7	739	4	B1931318	EST745259
37	940	58.1	739	5	BQ979990	CK282537
38	932.5	57.6	727	7	CN143032	CK282537
39	931.5	57.5	720	6	CD432672	CK282537
40	925	57.1	836	7	CK297749	CK282537
41	920.5	56.9	760	7	CF430522	CK282537
42	912	56.3	673	1	AL499638	CK282537
43	909	56.1	829	7	CK291957	CK282537
44	907.5	56.1	702	4	B1416734	CK282537
45	906	56.0	655	7	CK758916	CK282537

ALIGNMENTS

RESULT 1	894 bp	mrna	linear	EST 11-FEB-2004
BX839165	Arabidopsis thaliana	Adult vegetative tissue Col-0		
LOCUS	Arabidopsis thaliana	cDNA clone GSLTSL582B09	SPRIM, mRNA	sequence.
DEFINITION	BX839165	GI:42533248		
ACCESSION	BX839165	GI:42533248		
VERSION	BX839165	GI:42533248		
KEYWORDS	EST.			
SOURCE	Arabidopsis thaliana (thale cress)			
ORGANISM	Arabidopsis thaliana			
REFERENCE	Castelli, V., Aury, J.M., Jaillon, O., Wincker, P., Clepet, C., Menard, M., Cruaud, C., Quetier, F., Scarpetti, C., Schachter, V., Temple, G., Caboche, M., Weissenbach, J., and Salanoubat, M.			
AUTHORS	Whole Genome Sequences			
TITLE	A Combined Approach to Evaluate and Improve Arabidopsis Genome Annotation			
JOURNAL	Unpublished (2004)			
COMMENT	Genoscope - Centre National de Sequencage 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE Email: seqsfr@genoscope.cns.fr, Web: www.genoscope.cns.fr The sequences are based on single pass reads. Life Technologies (a division of Invitrogen) members carried out full-length libraries construction : Temple G. Genoscope members carried out sequencing and annotation : Castelli V., Aury J.M., Jaillon O., Wincker P., Menard M., Cruaud C., Schachter V., Weissenbach J., Salanoubat M. URGV INRA : Clepet C., Caboche M. Annotation is based on the June 2003 version of the Arabidopsis			

genome released by MIPS (Munich Information center for Protein Sequences).
http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/EST
<http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis>.

FEATURES

source

1..894
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /ecotype="Col-0"
 /db_xref="taxon:3702"
 /clone="GSLTSL582B09"
 /tissue_type="Adult vegetative tissue"
 /clone_lib="Arabidopsis thaliana Adult vegetative tissue"
 Col-0"

ORIGIN

Alignment Scores:
 Pred. No.: 1,12e-139 Length: 894
 Score: 1367.00 Matches: 272
 Percent Similarity: 97.56% Conservatives: 8
 Best Local Similarity: 94.77% Mismatches: 6
 Query Match: 84.43% Indels: 1
 DB: 5 Gaps: 0

US-10-606-300-11 (1-319) x BX839165 (1-894)

Qy 1 MetThrLysIleGluLeuArgAlaLeuGlyAenThrGlyLeuLysValSerAlaValGly 20
 Db 34 ATGCGAAATAGAGCTTCGAGCTTGGGGAACACAGGGCTTAAGGTTAGCGCGCTTGGT 93
 Qy 21 PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGlu-AspAspAlaValAl 40
 Db 94 TTGGTGCTCTCGCTCGAAGTGTCTTCGGTCAGTCGCGAAGGATGATGCGCTGCG 153
 Qy 40 aThrValArgGluAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyrTyrG1 60
 Db 154 CACGTCGCGAGGCTTCGCTCGGTATCACTTCTTCACACCTCCCGTATTATGG 213
 Qy 60 yGlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSe 80
 Db 214 AGGAACACTGTCTGAGAAATGCTTGGTAAGGACTTAAGGCTTTCGAAGTCCCTAGAAG 273
 Qy 80 rAspTyrIleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaG1 100
 Db 274 TGACTACATTGGCTACTAAGTGGTAGATATGGAGGAGGTTTGATTTTCAGTGCTGA 333
 Qy 100 uArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAsp11 120
 Db 334 CAGAGTAAGAGAGATATTGACGAGAGCTTCGAGAGGCTTCAGCTTGATTTATGTCAT 393
 Qy 120 eLeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePr 140
 Db 394 ACTTCATTGCCATGACATTGAGTTGGGTCTCTTGATCAGATTGTGAGTGAGACAAATCC 453
 Qy 140 oAlaLeuGlnLysLeuLysGlnGluGlyValThrArgPheIleGlyIleThrGlyLeuPr 160
 Db 454 TGCTCTTCAGAGACTGAACACGAGAGGGGAAACCGGTTCACTGATATCACTGCTTCC 513
 Qy 160 oLeuAspIlePheThrTyrValLeuAspArgValProGlyThrValAspValIleLe 180
 Db 514 GTTAGATTATTAATTATGTTATGATCGAGTGCCTCAGGAGCTGTCGATGTGATATT 573
 Qy 180 uSerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLy 200
 Db 574 GTCACTACTGTCAATACGGCGTTGATGATTGACGCTTGTGGAATTTACTTACTTACTTGA 633
 Qy 200 sSerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluG1 220
 Db 634 GAGCAAAAGGTGGGTGTGAAGTGTCTTCCATTAGCAATGGGCTTCCCTTACAGACA 693
 Qy 220 nGlyProGluTyrHisProAlaSerProGluLeuLysSerAlaSerLysAlaAlaVa 240
 Db 694 AGGTCTCTCAATGGCAACCTGCTTCCCTGAGCTCAAGTCTCGGAGCGAAGCGCGAGT 753

Qy 240 lAlaHisCysLysSerLysGlyLysLysIleThrLysLeuAlaLeuGlnTyrSerLeuAl 260
 Db 754 TGCTCACTGCAAAATGAGAGGGCAAGAGATCACAAAGTTAGCTCTGCAATACAGTTTAC 813
 Qy 260 aAsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGluAs 280
 Db 814 ANACGAGGAGATTTCGTCGGTGGTGGATGAGATATGCTCACAGGTAGAAAGAGAA 873
 Qy 280 nValAlaAlaValThrGlu 286
 Db 874 TGTTGCAGCACTTACAGAG 892

RESULT 2

CK253464

LOCUS

DEFINITION

CK253464

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Solanum tuberosum (potato)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Robin Buell

The Institute for Genomic Research

9712 Medical Center Dr, Rockville, MD 20850, USA

Email: potato-array@tigr.org

Clones can be requested from the University of Arizona Genomics

Institute via <http://genome.arizona.edu/orders/>

Seq primer: ATT TAG GTG ACA CTA TAG.

Location/Qualifiers

1..953

/organism="Solanum tuberosum"

/mol_type="mRNA"

/cultivar="Kennebec"

/db_xref="taxon:4113"

/clones="POCC562"

/tissue_type="callus"

/lab_host="DH10B-Tona"

full-length"

/clone_lib="potato callus cDNA library, normalized and

/note="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;

supplier: RNA was isolated from Solanum tuberosum var.

Kennebec callus tissue grown on solid media."

ORIGIN

Alignment Scores:

Pred. No.: 3,91e-133 Length: 953

Score: 1308.00 Matches: 253

Percent Similarity: 88.22% Conservatives: 24

Best Local Similarity: 80.57% Mismatches: 37

Query Match: 80.79% Indels: 0

DB: 7 Gaps: 0

US-10-606-300-11 (1-319) x CK253464 (1-953)

Qy 4 IleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGlyPheGlyAla 23

Db 10 TTGCAGCTCCGACCACTGGGCAACACTGAGCTCAATCTCAGCTCTGTCGGTTTCGGCGT 69

Qy 24 SerProLeuGlySerValPheGlyProValAlaGluAspAlaValAlaThrValArg 43

Db 70 TCTCCTTCGCAAGGTTTCGGCGATGTCCTCCAAACAGACGCTTCGCCCGCGCTCCGG 129

Qy 44 GluAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyrTyrGlyThrLeu 63


```

Db      685 GTGGGGGATCAGTCTCTCTCTTCAATGGGTCTTCTTACAGCGCTGGTGGTCCA 744
QY      224 GluTrpHisProAlaSerProGluLeuLysSerAlaSerLysAlaAlaValAlaHisCys 243
Db      745 GAGTGGCACCTGCTTCTTCTGAACCTTAAGGCTGCTCCGAGCTGCTGCTGATCACTGC 804
QY      244 LysSerLysGlyLysGlyLysLeuAlaLeuGlnThrLysLeuAlaLeuAlaSerLysGlu 263
Db      805 AAAGGAAAGGGAAGAATATCTCAAAATTTAGCTTCCAGTACAGCTTAGCAAAATACCGAT 864
QY      264 IleSerSerValLeuValGlyMetSerValSerGlnValGluGluAsnValAlaAla 283
Db      865 ATTCTTCCATACATAGTGGGATGAAGTCTGTAAAGAGGTGGAGGAAATATAGCAGCT 924
QY      284 ValThrGluLeuGluSerLeuGlyMetAspGlnGlnThrLysLeuGluValGluAlaIle 303
Db      925 GCCCTAGACTAGCAACGGCTGGATGGATGGATGAAGAAATATATACAGATCACACATT 984
QY      304 LeuGluPro 306
Db      985 CTGGAACCC 993

```

RESULT 4
CK250147
LOCUS
DEFINITION
EST733784 potato callus cDNA clone POCBD80 5' end, mRNA sequence.

ACCESSION
CK250147
VERSION
CK250147.1 GI:39801934

KEYWORDS
SOURCE
Solanum tuberosum (potato)

ORGANISM
Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum.

1 (bases 1 to 987)

Buell, C.R., Hart, A., Zismann, V., Karamycheva, S.A. and Baker, B.

Generation of ESTs from potato callus tissue

Unpublished (2003)

Other ESTs: EST733785

Contact: Robin Buell

The Institute for Genomic Research

9712 Medical Center Dr, Rockville, MD 20850, USA

Email: potato-array@tigr.org

Clones can be requested from the University of Arizona Genomics

Institute via http://genome.arizona.edu/orders/

Seq primer: ATT TAG GTG ACA CTA TAG.

Location/Qualifiers

1..987

/organism="Solanum tuberosum"

/mol_type="mRNA"

/cultivar="Kennebec"

/db_xref="taxon:4113"

/clone="POCBD80"

/tissue_type="callus"

/lab_host="DH10B-TonA"

/clone_lib="potato callus cDNA library, normalized and

full-length"

/note="Vector: pCMVSPORT5.1; Site 1: EcoRI; Site 2: NotI;

supplier: RNA was isolated from Solanum tuberosum var.

Kennebec callus tissue grown on solid media."

ORIGIN

Alignment Scores:

Pred. No.:	2,79e-126	Length:	987
Score:	1246.00	Matches:	242
Percent Similarity:	88.33%	Conservative:	23
Best Local Similarity:	80.67%	Mismatches:	35
Query Match:	76.96%	Indels:	0
DB:	7	Gaps:	0

US-10-606-300-11 (1-319) x CK250147 (1-987)

```

QY      4   IleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValAlGlyPheGlyAla 23
Db      87   TTGAGCTCCGATCAGTCTGGGCAACACACAGTCAATCTCAGCTCTGTGGCTTCGGCGCT 146
QY      24   SerProLeuGlySerValPheGlyProValAlaAGluAspAspAlaValAlaThrValArg 43
Db      147  TCTCTCTTCGGCAAGGTTTTCGGCGATGCTCCGGAACAAGACGCTTCGCCGCCGCTCCGG 206
QY      44   GluAlaPheArgLeuGlyLysPheAspThrSerProTyrTyrGlyLysThrLeu 63
Db      207  GAGGCTTTCGCTCGGCGTCAATTTCTTGACACTTCCCGTATTAATGGAGGAACGTTA 266
QY      64   SerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProAspSerAspTyrIle 83
Db      267  TCGGAAAAGGTAAGGCTTGAAGGCTCTTGAGAGTACCTTAGAGATCAGTACATTT 326
QY      84   ValAlaThrLysCysGlyArgTyrLysGlyPheAspPheSerAlaGluArgValArg 103
Db      327  GTGCTCACAAATGTGGAGGTACAAAGAGGATTTGAATTTTAGTGTGAGAGAGTGACT 386
QY      104  LysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIleLeuHisCys 123
Db      387  AAAAGTATTGATGAGAGCTTCGAGAGGCTCGAGCTTGATTATGTTGATATGTTACATGT 446
QY      124  HisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIleProAlaLeuGln 143
Db      447  CATGATATTGAATTTGGGTCGCTCGATCAGATTGTGAATGAGACACTTCCCGCCCTTCTA 506
QY      144  LysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGlyLeuProLeuAspIle 163
Db      507  AAATTTGAAGCAAGCTGGAAAGATCCGCTTCATTGGTATAACCGGCTACCTTTGGGGATA 566
QY      164  PheThrTyrValLeuAspArgValProGlyThrValAspValIleLeuSerTyrCys 183
Db      567  TTCATTATGTTGCTTGATCGGTCCTCCAGGCACAGTTGATGTCATCTGTCATATTGT 626
QY      184  HisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLysSerLysGly 203
Db      627  CACTACTGTGTCAACGATTCAACTTTGAGGATCTGTGCCATACCTGAAGAGCAAGGT 686
QY      204  ValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGlnGlyProPro 223
Db      687  GTGGGGGTGATCAGTGCTTCCTCTTCAATGGGTCTCTTACAGAGCTGGTGGTCCA 746
QY      224  GluTrpHisProAlaSerProGluLeuLysSerAlaSerLysAlaAlaValAlaHisCys 243
Db      747  GAGTGGCACCTGCTTCTTCTGAACTTAAGGCTGCTCCGAGCTGCTGTCATCACTGC 806
QY      244  LysSerLysGlyLysLysIleThrLysLeuAlaLeuGlnThrSerLeuAlaAsnLysGlu 263
Db      807  AAAGGAAAGGGAAGAATATCTCAAAATTTAGCCTTTCAGCTAGCTTACAGTAAATACCGAT 866
QY      264  IleSerSerValLeuValGlyMetSerSerValSerGlnValGluGluAsnValAlaAla 283
Db      867  ATTCTTCCATACATAGTGGGATGAAGTCTGTAAAGAGGTGGAGGAAATATATAGCAGCT 926
QY      284  ValThrGluLeuGluSerLeuGlyMetAspGlnGlnThrLysLeuGluValGluAlaIle 303
Db      927  GCCCTAGACTAGCAACGGCTGGGATGGATGAAGAAACATTATATCAGAGATCAGACATT 986

```

RESULT 5

CK243867

LOCUS

DEFINITION

EST727504 potato callus cDNA library, normalized and full-length

Solanum tuberosum cDNA clone POC594 5' end, mRNA sequence.

ACCESSION

CK243867

VERSION

CK243867.1

GI:39788889

KEYWORDS

EST.

SOURCE

ORGANISM

Solanum tuberosum (potato)

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

asterids; lamids; Solanales; Solanaceae; Solanum.
 1 (bases 1 to 948)
 REFERENCE Buell, C.R., Hart, A., Zismann, V., Karamycheva, S.A. and Baker, B.
 AUTHORS Generation of ESTs from potato callus tissue
 TITLE Unpublished (2003)
 JOURNAL
 COMMENT Other ESTs: EST727503
 Contact: Robin Buell
 The Institute for Genomic Research
 9712 Medical Center Dr, Rockville, MD 20850, USA
 Email: potato-array@tigr.org
 Clones can be requested from the University of Arizona Genomics
 Institute via http://genome.arizona.edu/orders/
 Seq primer: ATT TAG GTG ACA CTA TAG.

FEATURES

Location/Qualifiers
 1..948
 /organism="Solanum tuberosum"
 /mol_type="mRNA"
 /cultivar="Kennebec"
 /db_xref="taxon:4113"
 /clone="POCA594"
 /tissue_type="callus"
 /lab_host="DH10B-Tona"
 /clone_lib="potato callus cDNA library, normalized and full-length"
 /note="vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI; supplier: RNA was isolated from Solanum tuberosum var. Kennebec callus tissue grown on solid media."

ORIGIN

Alignment Scores:
 Pred. No.: 9.16e-125 Length: 948
 Score: 1232.00 Matches: 241
 Percent Similarity: 88.59% Conservative: 23
 Best Local Similarity: 80.87% Mismatches: 33
 Query Match: 76.10% Indels: 1
 DB: 7 Gaps: 0

US-10-606-300-11 (1-319) x CK243867 (1-948)

Qy 4 IleGluLeuArgAlaLeuGlyAsnThrGlyLeuValSerAlaValGlyPheGlyVala 23
 Db 52 TTGCAGCTCCGATCACTCGGCAACACTGGATCTCAATCTCAGCTCTGTCGGCTTCGGCGCT 111
 Qy 24 SerProLeuGlySerValPheGlyProValAlaGluAspAlaValAlaThrValArg 43
 Db 112 TCTCTCTCGGCAAGTTTTCGGCGATCTCCGACACAGCGCTTCGGCGCGCTCGG 171
 Qy 44 GluAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyrTyrGlyThrLeu 63
 Db 172 GAAGCTTTTCGCTCGCGCTCAATTTCTTTCGACACTTCGCGTATTATGGAGAACGTTA 231
 Qy 64 SerGluLeuMetLeuGlyLeuGlyLeuValAlaGluInValProArgSerAspTyrile 83
 Db 232 TCGGAAAGGTTACTAGGAAAGCTTTGAAGGCTCTTGGAGTACCTTAGAGATCAGTACAT 291
 Qy 84 ValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGluArgValArg 103
 Db 292 GTGCTACAAATGTGGAGGTACAAAGGGATTGATTTTGTGCTGAGAGTGTACT 351
 Qy 104 LysSerIleAspGluSerLeuGluArgLeuGluInLeuAspTyrValAlaPheLeuHisCys 123
 Db 352 AAAAGTATTGATGAGAGCTTGGAGAGGCTTCGAGCTTGATATTGTTGATGTTACATGT 411
 Qy 124 HisApIleGluPheGlySerLeuAspGlnIleValSerGluThrIleProAlaLeuGln 143
 Db 412 CATGATATTGAATTTGGGTCCCTCGATCAGATTGTGAATGAGACACTTCCCGCCCTCTTA 471
 Qy 144 LysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGlyLeuProLeuAspIle 163
 Db 472 AAATTGAAGCAGCTGGAAAGATCCGCTTCATTTGGTATACCGGCTACCTTTGGGGATA 531
 Qy 164 PheThrTyrValLeuAspArgValProGlyThrValAspValIleLeuSerTyrCys 183

Db 532 TTCATTATATGTGTGATCGCGTCCCTCCAGGCACAGTTGATGTCATCCTGTCTATATTGT 591
 Qy 184 HisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLysSerLysGly 203
 Db 592 CACTACTGTGTCAACGATTCAACTTTGGAGGATCTGTTGCCATACCTACCTGAAGCAAGGT 651
 Qy 204 ValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGlnGlyProPro 223
 Db 652 GTGGGGGTGATCAGTCTCTCTCTTCAATGGGTCTTCTTACTGAGGCTGGTGTGCTCA 711
 Qy 224 GluThrHisProAlaSerProGluLeuLysSerAlaSerLysAlaAlaValAlaHisCys 243
 Db 712 GAGTGGCACCTCTCTTCTTGAACCTTAAAGCTGCTCCGAGCTGCTGCTGATCAGTGC 771
 Qy 244 LysSerLysGlyLysLysIleThrLysLeuAlaLeuGluInTyrSerLeuAlaAsnLysGlu 263
 Db 772 AAAGAAAGGGAAGAATAATCTCAAAATTAGCTTGCAGTACAGCTTAGCAAAATACCGAT 831
 Qy 264 IleSerSerValLeuVal-GlyMetSerSerValSerGlnValGluGluAsnValAlaAl 283
 Db 832 ATTTCTTCCATCTAGTGGGGGATCAATCTGTTAAAGAGCTGGAGGAAATATAGCAGC 891
 Qy 283 aValThrGluLeuGluSerLeuGlyMetAspGluGluThrLeuSerGluVal 300
 Db 892 TGCCTAGAACTAGCAACGCTGGATGGATGAAGAACAATATCAGAGATC 943

RESULT 6
 CK282536 979 bp mRNA linear EST 02-AUG-2004
 LOCUS EST745258 Nicotiana benthamiana mixed tissue cDNA library, NMAD20 5'
 DEFINITION normalized, full-length Nicotiana benthamiana cDNA clone NMAD20 5'
 end, mRNA sequence.
 ACCESSION CK282536.1 GI:39854240
 VERSION CK282536
 KEYWORDS EST.
 SOURCE Nicotiana benthamiana
 ORGANISM Nicotiana benthamiana
 REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamids; Solanales; Solanaceae; Nicotiana.
 1 (bases 1 to 979)
 Buell, C.R., Hart, A., Zismann, V., Karamycheva, S.A., Day, B.,
 Skaskawicz, B., Jin, H. and Baker, B.
 Generation of EST sequences from Nicotiana benthamiana
 Unpublished (2003)
 Other ESTs: EST745256 EST745257 EST745259
 Contact: Robin Buell
 The Institute for Genomic Research
 9712 Medical Center Dr, Rockville, MD 20850, USA
 Email: potato-array@tigr.org
 Clones can be requested from the University of Arizona Genomics
 Institute via http://genome.arizona.edu/orders/
 Seq primer: ATT TAG GTG ACA CTA TAG.

FEATURES

Location/Qualifiers
 1..979
 /organism="Nicotiana benthamiana"
 /mol_type="mRNA"
 /db_xref="taxon:4100"
 /clone="NMAD20"
 /tissue_type="abiotic and biotic stress-treated leaves, callus tissue and root tissue"
 /lab_host="DH10B-Tona"
 /clone_lib="Nicotiana benthamiana mixed tissue cDNA library, normalized, full-length"
 /note="vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI; supplier: RNA was isolated from Nicotiana benthamiana tissues that include callus, roots from liquid culture grown plants, heat-stressed leaves (38 C, 3 hr and 6 hr), cold-stressed leaves (5 C 3 hr, 6hr), and pathogen challenged leaves (Pseudomonas syringae pv tomato 12 hr; Xanthomonas campestris pv campestris 12 hr, 18hr; Pseudomonas syringae pv phaseolicola 18hr, and Xanthomonas campestris pv vesicatoria 18hr). RNA was isolated from

QY 84 ValAlaThrLysCysGlyArgTyrLys---GluGlyPheAspPheSerAlaGluArgVal 102
 Db 324 GTCGCCCAACAGTGGCGCGCTACAAAGACGAGGGTTTCGATCTTCGCGCAACCGGTGTG 383
 QY 103 ArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIleLeuHis 122
 Db 384 ACACCCAGCATAGACAGAGCTCCGCCACTGGGGCTGGATGATGATGATGATGATGATGAT 443
 QY 123 CysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIleProAlaLeu 142
 Db 444 GCTCATGACATCGAGTTCACCCATCTCGAACAGATGTTGATGATGATGATGATGATGAT 503
 QY 143 GlnLysLeuLysGlnGluGlyThrArgPheIleGlyIleThrGlyLeuProLeuAsp 162
 Db 504 CAGAAAGTCAAGGAGAAATGGGAAGCAGCGTTTCATTGGCATCACCAGGTTGCTCTCAGC 563
 QY 163 IlePheThrTyrValLeuAspArgValProGlyThrValAspValIleLeuSerTyr 182
 Db 564 ATCTACCTTATGCTCTGACCGTGTAGCACCGAGCTCGGTGGAGTGTATCTATCTTAT 623
 QY 183 CysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLysSerLys 202
 Db 624 TGTCACTACGGGATCAATGACACCTCCCTTGTGATCTGCTCCCTACTTGAAGACAA 683
 QY 203 GlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGlnGlyPro 222
 Db 684 GGTGTGGGGTATCAGTGCTTCGCCCTCTCCATGGGTCTTTTAAACAGATAATGGGCA 743
 QY 223 ProGluTyrHisProAlaSerProGluLysSerAlaSerLysAlaAlaValAlaHis 242
 Db 744 CCGGAGTGGACCCCTGACACCAAGAACTTAAGTCAGCATCGAAGGCTGCAGCAGATCAC 803
 QY 243 CysLysSerLysGlyLysLysIleThrLysLeuAlaLeuGlnTyrSerLeuAlaAsnLys 262
 Db 804 TGTAGAAGAGGGGAAAGCAATACAGAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 863
 QY 263 GluIleSerValLeuValGlyMetSerSerValSerGlnValGluGluAsnValAla 282
 Db 864 GAAATCTCGACAGTTCCTGTTGGATGAATCACTTTTGGAAACAGGTGGAGGAGAAATGGCG 923
 QY 283 AlaValThrGluLeuSerLeuGlyMetAspGlnGluThrLeuSerGluValGluAla 302
 Db 924 GCTGCACCTGGAGTGTGTCAACATCAGGCATCGATGAAGAACTTATGCGTGAAGTTGAAGCA 983
 QY 303 IleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGln 318
 Db 984 ATTCTGAGCCTGTGAGAAATCTGACATGGCTAGCGGCATCCAAACAA 1031

RESULT 8
 CL979301
 LOCUS
 DEFINITION
 OsIFCC033028 Oryza sativa Expressed Sequence Tag (EST) Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

OsIFCC033028 Oryza sativa (indica cultivar-group)
 Oryza sativa (indica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzoideae; Oryza.

1 (bases 1 to 1302)
 Ma L., Wang J., Chen C., Liu X., Su N., Li L., Wang X., Cao M., Jiao Y., Sun N., Zhang X., Bao J., Sun D., Zhao H., Yuan L., Wong G.K.S., Deng X.W. and Wang J.

An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis
 Unpublished (2004)
 CONTACT: Chen Chen
 Department of Bioinformatics
 Beijing Institute of Genomics
 Chinese Academy of Sciences, Beijing 101300, China
 Tel: 86-10-80481559

Fax: 86-10-80488676
 Email: chenchen@genomics.org.cn
 Rice genomic sequence.
 Class: exon-trapped.

FEATURES

source
 1..1302
 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="genomic DNA"
 /db_xref="taxon:39946"
 /clone_lib="Oryza sativa Expressed Library"
 /note="Oryza sativa exon trapped genomic sequences"

ORIGIN

Alignment Scores:
 Pred. No.: 8,3e-122 Length: 1302
 Score: 1207.00 Matches: 235
 Percent Similarity: 82.24% Conservative: 29
 Best Local Similarity: 73.21% Mismatches: 51
 Query Match: 74.55% Indels: 6
 Gaps: 1

US-10-606-300-11 (1-319) x CL979301 (1-1302)

QY 4 IleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGlyPheGlyAla 23
 Db 334 ATGGAGCTCCGCGAGCTCGCGCCACGCGCTCCGCGTCAGCCCGTCCGGTTCGGCGCC 393
 QY 24 SerProLeuGlySerValPheGlyProValAlaGluAspAspAlaValAlaThrValArg 43
 Db 394 TCCCTCTCGGCCATGTTCTTCGGTGACGTCCCGCGAGCTCCCGCGCGCGCGCGCTCCGC 453
 QY 44 GluAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyrTyrGlyThrLeu 63
 Db 454 CGCGCCCTCGACCTCGGCATCACTTCTTCGACACCTCCCGCTACTACGCGGCGACGGTG 513
 QY 64 SerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAspTyrIle 83
 Db 514 TCGGAGTCCGTCGTCGCGGAGTGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGTG 573
 QY 84 ValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGluArgValArg 103
 Db 574 GTCGCCCAACAGTGGCGCGCTACAGGAAGGGTTTCGACTTCAGCGCGCGCGCGCGCTCACC 633
 QY 104 LysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIleLeuHisCys 123
 Db 634 CGCAGCTCGACGAGAGCTCGACGCTCGCGCTCGACTACGTCAGTCGACATCTCTCCACTGC 693
 QY 124 HisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIleProAlaLeuGln 143
 Db 694 CACGACATCGAGTTCACCGACCTCGACCATGTAATGAGACGATTCGCGTGTCTCCAG 753
 QY 144 LysLeuLysGlnGluGlyThrArgPheIleGlyIleThrGlyLeuProLeuAspIle 163
 Db 754 AAGATCAAGGAGAGCGCGGAGCGCGGTTCATCGGGATAACCGGCGCTCCCATTTGAGCATC 813
 QY 164 PheThrTyrValLeuAspArgValProGlyThrValAspValIleLeuSerTyrCys 183
 Db 814 TACACTTATGTCTCGACCGGCTCGCGCGAGTTCGTGGATGTGATCTGTCTTACTTGC 873
 QY 184 HisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLysSerLysGly 203
 Db 874 CACTATGGGATCAATGATACCGCGCTGTGGATTTGCTTCCCTACCTCAAGAGCAAGGT 933
 QY 204 ValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGlnGlyProPro 223
 Db 934 GTTCGCGTGTATGATGCTTCACCTCTTGCATATGGGCTTCTTACAGATAAGCGGCGCC 993
 QY 224 GluTrpHisProAlaSerProGluLysSerAlaSerLysAlaAlaValAlaHisCys 243
 Db 994 GAATGGCACCTCTGCACCGAAGAACTTAAGTTGGATGTCAGCGGCGAGCAGATCCTGT 1053
 QY 244 LysSerLysGlyLysLysIleThrLysLeuAlaLeuGlnTyrSerLeuAlaAsnLysGlu 263

Db 475 CATGATATTGAGTTGGTCTCTTGATCAGATTGTGAATGAGACGCTCCAGCCCTTCAG 534

Qy 144 LysLeuLysGlnGluGlyThrArgPheIleGlyIleThrGlyLeuProLeuAspIle 163

Db 535 AAACCTGAAGCAAGCTGGAAAGATCGTTTCATTGGTATATAACCGGCTTCCATTGGGGATA 594

Qy 164 PheThrTyrValLeuAspArgValProGlyThrValAspValIleLeuSerTyrCys 183

Db 595 TTCACCTTATGCTTGATCGGGTACCTCCAGGAACGGTTGATGTATTCTGTCTATATTGT 654

Qy 184 HisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLysSerLysGly 203

Db 655 CACTACAGTATCAATGACTCACTTGGAGATCGGTGGCCATACCTGGAAGACAGGCT 714

Qy 204 ValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGlnGlnProPro 223

Db 715 GTGGAGGTGATCAGTCTTCCCTCTTGCAATGGTCTTCTTACTGAGGCTGGAGCTCCA 774

Qy 224 GluTrpHisProAlaSerProGluLeuLysSerAlaSerLysAlaValAlaHisCys 243

Db 775 GAGTGGCAACCTGCTTCTTGAACCTTAAGCTGCTCCGAGCTGCGGTTCATTCATTCG 834

Qy 244 LysSerLysGlyValLysIleThrLysLeuAlaLeuGlnTyrSerLeuAlaAsnLysGlu 263

Db 835 AAAGAAAGGAAAGATATATCAAAATAGCTTTGCATACAGTTTACGAAATACCGAT 894

Qy 264 IleSerSerValLeuValGlyMet 271

Db 895 ATTTTCATCCGTACTAGTGGGGATG 918

RESULT 13

CK286441

LOCUS

DEFINITION EST749163 Nicotiana benthamiana mixed tissue cDNA library, normalized, full-length Nicotiana benthamiana cDNA clone NMB212 5' end, mRNA sequence.

ACCESSION CK286441

VERSION CK286441.1 GI:39862007

KEYWORDS

SOURCE Nicotiana benthamiana

ORGANISM Nicotiana benthamiana

REFERENCE 1 (bases 1 to 896)

AUTHORS Buehl, C.R., Hart, A., Ziesmann, V., Karamycheva, S.A., Day, B., Staskawicz, B., Jin, H. and Baker, B.

TITLE Generation of EST sequences from Nicotiana benthamiana

JOURNAL Unpublished (2003)

COMMENT Other ESTs: EST749164

Contact: Robin Buehl

The Institute for Genomic Research

9712 Medical Center Dr, Rockville, MD 20850, USA

Email: potato-array@tigr.org

Clones can be requested from the University of Arizona Genomics Institute via <http://genome.arizona.edu/orders/>.

Seq primer: ATT TAG GTG ACA CTA TAG.

FEATURES

source

1..896

/organism="Nicotiana benthamiana"

/mol_type="mRNA"

/db_xref="taxon:4100"

/clones="NMB212"

/tissue_type="abiotic and biotic stress-treated leaves, callus tissue and root tissue"

/lab_hosts="DH10B-Ton"

/clone_lib="Nicotiana benthamiana mixed tissue cDNA library, normalized, full-length"

/note="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI; supplier: RNA was isolated from Nicotiana benthamiana tissues that include callus, roots from liquid culture grown plants, heat-stressed leaves (38 C, 3 hr and 6 hr), cold-stressed leaves (5 C 3 hr, 6hr), and pathogen

challenged leaves (*Pseudomonas syringae* pv tomato 12 hr; *Xanthomonas campestris* pv campestris 12 hr, 18hr; *Pseudomonas syringae* pv phaseolicola 18hr, and *Xanthomonas campestris* pv vesicatoria 18hr). RNA was isolated from these tissues and pooled in approximately equal molar amounts."

ORIGIN

Alignment Scores:

Pred. No.: 1,84e-112 Length: 896

Score: 1120.00 Matches: 216

Percent Similarity: 89.88% Conservat: 15

Best Local Similarity: 84.95% Mismatches: 26

Query Match: 69.18% Indels: 0

DB: 7 Gaps: 0

US-10-606-300-11 (1-319) x CK286441 (1-896)

Qy 4 IleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValAlaGlyPheGlyAla 23

Db 124 TTGCAGCTCCGACCACTGGGCAACACCGGCTCAACCTTAGCTCCGTCGGCTTCGGCGCT 183

Qy 24 SerProLeuGlySerValPheGlyProValAlaGluAspAspAlaValAlaThrValArg 43

Db 184 TCTCTCTTGGCAAGTCTTCGGCGATGCTCCGAAACAGACGCTTCGCCGCCGTGGCGC 243

Qy 44 GluAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyrTyrGlyThrLeu 63

Db 244 GAAGCCTTTCGCTTGGGTCATTTCTTGATACTTCCCGTATTATGGAGGAACGTTA 303

Qy 64 SerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAspTyrIle 83

Db 304 TCAGAAAAGGTACTAGGGAAGGCTTTGAAGGCTCTTGGAGCTCTCTAGAGATGAGTACAT 363

Qy 84 ValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGluArgValArg 103

Db 364 GTGTCAACAAAGTGTGGAGGTACAAGAGAGGATTTGATTTTCAGTGTGAGAGATGACT 423

Qy 104 LysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIleLeuHisCys 123

Db 424 AAAGCATTTGATGAGAGCTGGAGAGGCTACAGCTTGATTTGATATTTTGAATGT 483

Qy 124 HisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIleProAlaLeuGln 143

Db 484 CATGATATTGAGTTTGGGTCTCTTGTATCAGATTGTGAATGAGACGCTTCCAGCCCTTCAG 543

Qy 144 LysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGlyLeuProLeuAspIle 163

Db 544 AAACCTGAAGCAAGCTGGAAAGATCGTTTCATTGGTATACCGGCTTCCATTGGGGATA 603

Qy 164 PheThrTyrValLeuAspArgValProGlyThrValAspValIleLeuSerTyrCys 183

Db 604 TTCACCTTATGCTTGATCGGGTACCTCCAGGAACGGTTGATGTATTCTGTCTATATTGT 663

Qy 184 HisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLysSerLysGly 203

Db 664 CACTACAGTATCAATGACTCACTTGGAGGATCGGTTCGCCATACCTGAAGAGCAAGGCT 723

Qy 204 ValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGlnGlnProPro 223

Db 724 GTGGAGGTGATCAGTCTTCCCTCTTGCAATGGTCTTCTTACTGAGGCTGGAGCTCCA 783

Qy 224 GluTrpHisProAlaSerProGluLeuLysSerAlaSerLysAlaValAlaHisCys 243

Db 784 GAGTGGCAACCTGCTTCTTGAACCTTAAGGCTGCTCCGAGCTGCGGTTCATTCATTCG 843

Qy 244 LysSerLysGlyLysLysIleThrLysLeuAlaLeuGlnTyrSerLeuAla 260

Db 844 AAAGAAAGGAAAGATATATCAAAATAGCTTTGCAATACAGTTTAGCA 894

RESULT 14

CK282535/c

LOCUS

926 bp mRNA linear EST 02-AUG-2004

QY	138	ThrIleProAlaLeuGlnLysLeuLysGlnGluGlyValArgPheIleGlyThr	157
Db	684	AGCTTCCAGCCCTTCAGAACAACCTGAAGCAAGCTGGAAAGATCGTTTTCATTGTGATAAACC	625
QY	158	GlyLeuProLeuAspIlePheThrTyrrValLeuAspArgValProProGlyThrValAsp	177
Db	624	GGCCTCCATTGGGGATATTCATTATGTGCTTGATCGGGTACCTCCAGAACGGTTGAT	565
QY	178	ValIleLeuSerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuPro	197
Db	564	GTTATTCTGTGCATATTGTCACTACAGTACATCAATGACTCAACTTTGGAGGATCCGTTGCCA	505
QY	198	TyrLeuLysSerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeu	217
Db	504	TACCTGAAGAGCAGGGGTGTGGAGTGATCAGTGCTTCCCCTCTTGCATATGGGTCTTCTT	445
QY	218	ThrGluGlnGlyProProGluTrpHisProAlaSerProGluLeuLysSerAlaSerLys	237
Db	444	ACTGAGGCTGGAGCTCCAGATGGCACCCCTGCTTCTTCTGAACCTTAAGGCTGCTGCCGA	385
QY	238	AlaAlaValAlaHisCysLysSerLysGlyLysLysIleThrLysLeuAlaLeuGlnTyr	257
Db	384	GCTGCCGTTGATCATTTGCCAAGGAAGGAAGAAATATATCAAATATAGCCTTGCATATC	325
QY	258	SerLeuAlaAsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnVal	277
Db	324	AGTTTAGCAATATACCATATTTCATCCGTACTAGTGGGATGAATCCGTTAAAGAGGTG	265
QY	278	GluGluAsnValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeu	297
Db	264	GAGGAAAATATAGGAGCTGCCCTGGAGCTAGCAACGGCTGGGATGGATGAAGAACTTTA	205
QY	298	SerGluValGluAlaIleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHis	317
Db	204	TCAGAGATCACAGACATTTCTGAACCCAGTGAAGATCAGTCAITGGCTAGTGTATCCAA	145
QY	318	GlnAsn	319
Db	144	CNAAGT	139
RESULT	15		
CK297635			
LOCUS	EST760349	Nicotiana benthamiana mixed tissue cdna library, normalized, full-length Nicotiana benthamiana cDNA clone NEMDA27 5', end, mRNA sequence.	EST 02-AUG-2004
DEFINITION			
ACCESSION	CK297635		
VERSION	CK297635.1	GI:39884220	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS	Buell,C.R., Hart,A., Zismann,V., Karamycheva,S.A., Day,B., Staskawicz,B., Jin,H. and Baker,B. Generation of EST sequences from Nicotiana benthamiana Unpublished (2003)		
TITLE	The Institute for Genomic Research		
JOURNAL	9712 Medical Center Dr, Rockville, MD 20850, USA		
COMMENT	Email: potato-array@tigr.org Clones can be requested from the University of Arizona Genomics Institute via http://genome.arizona.edu/orders/ Seq primer: ATT TAG GTG ACA CTA TAG.		
FEATURES			
source	1..859	Location/Qualifiers	
	/organism="Nicotiana benthamiana"		
	/mol_type="mRNA"		
	/db_xref="taxon:4100"		
	/clone="NEMDA27"		

/tissue_type="abiotic and biotic stress-treated leaves,
callus tissue and root tissue"
/lab_host="DH10B-Tona"
/clone_lib="Nicotiana benthamiana mixed tissue cDNA
library, normalized, full-length"
/note="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
supplier: RNA was isolated from Nicotiana benthamiana
tissues that include callus, roots from liquid culture
grown plants, heat-stressed leaves (38 C, 3 hr and 6 hr),
cold-stressed leaves (5 C 3 hr, 6hr), and pathogen
challenged leaves (Pseudomonas syringae pv tomato 12 hr;
Xanthomonas campestris pv campestris 12 hr, 18hr;
Pseudomonas syringae pv phaseolicola 18hr, and Xanthomonas
campestris pv vesicatoria 18hr). RNA was isolated from
these tissues and pooled in approximately equal molar
amounts."

ORIGIN

Alignment Scores:

Pred. No.:	2.02e-107	Length:	859
Score:	1074.00	Matches:	206
Percent Similarity:	89.80%	Conservative:	14
Best Local Similarity:	84.08%	Mismatches:	25
Query Match:	56.34%	Indels:	0
DB:	7	Gaps:	0

US-10-606-300-11 (1-319) x CK297635 (1-859)

Qy	4	IleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGlyPheGlyAla	23
Db	118	TTGCAGCTCCGACCACTGGGCAACACCGGCTCAACCTAGCTCGCTGGGCTTGGGCGCT	177
Qy	24	SerProLeuGlySerValPheGlyProValAlaGluAspAlaValAlaThrValArg	43
Db	178	TCTCTCTTGGCAAGTCTTCGGCGATGTCCTCCGACCAAGACGCTTCGCCGCCGTGGCGC	237
Qy	44	GluAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyrTyrGlyGlyThrLeu	63
Db	238	GAAGCCTTTGGCGTCAATTTCTTGATCTTCCCGTATTATGGAGGAACGTGA	297
Qy	64	SerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAspTyrIle	83
Db	298	TCAGAAAAGGTACTAGGGAAGGCTTTGAAGGCTCTTGGAGCTCTTAGAGATGAGTACATT	357
Qy	84	ValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGluArgValArg	103
Db	358	GTGTCAACAAGTGTGGAGGTACAAAGAGGATTTGATTTTCAGTCTGAGAGAGTACT	417
Qy	104	LysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIleLeuHisCys	123
Db	418	AAAAGCATTTGATGAGAGCTTGGAGAGGCTACAGCTTGATTGATTGTTGATTTTGCATGT	477
Qy	124	HisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIleProAlaLeuGln	143
Db	478	CATGATATTGATTTGGGTCTCTTGATCAGATTGTGAATGAGACGCTTCCAGCCCTTCAG	537
Qy	144	LysLeuLysGlnGluGlyThrArgPheIleGlyIleThrGlyLeuProLeuAspIle	163
Db	538	AAACTGAGCAAGCTGGAAAGATCGTTTCATTGGTATTAACCGGCTTCCATTGGGGATA	597
Qy	164	PheThrTyrValLeuAspArgValProGlyThrValAspValIleLeuSerTyrCys	183
Db	598	TTCACTTATGCTTGTATCGGTACCTCCAGGAACGGTTGATGTTATTCTGTCTATATTGT	657
Qy	184	HisTyrGlyValAsnAspSerThrLeuLeuAspLeuProTyrTyrLeuLysSerLysGly	203
Db	658	CACACAGTATCAATGACTCACTTTGGAGGATCCGTTGCCATACCTGAAAGAGCAAGGGT	717
Qy	204	ValGlyValIleSerAlaSerProLeuAlaMetGlyLeuThrGluGlnGlyProPro	223
Db	718	GTGGAGGTGATCAGTGTCTCCCTCTTGCATGGGTCTTCTTACTGAGGCTGGAGCTCCA	777
Qy	224	GluTrpHisProAlaSerProGluLeuLysSerAlaSerLysAlaAlaValAlaHisCys	243

Db	778	GAGTGGCACCCCTGCTTCTTCTGAACCTTAAGGCTGCTCCGAGCTGCGTTGATCATTGC	837
Qy	244	LysSerLysGlyLys	248
Db	838	AAAGGAAGGGGAAG	852

Search completed: November 13, 2005, 11:14:51
Job time : 3695 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2005, 08:26:32 ; Search time 87 Seconds
(without alignments)
1418.121 Million cell updates/sec

Title: US-10-606-300-11

Perfect score: 1619

Sequence: 1 MTKIELRALGNTGLKVSAG.....VEAILEPVKNLTWPSGIHQN 319

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1619	100.0	319	4	Aae11998 Arabidops
2	1619	100.0	319	5	Aau76343 A. thalia
3	1583	97.8	357	3	Aag16520 Arabidops
4	1286	78.4	253	3	Aag16521 Arabidops
5	593.5	36.7	345	4	Abb67077 Drosophil
6	544.5	33.6	342	4	Abb71229 Drosophil
7	420.5	26.0	294	4	Abb60118 Drosophil
8	366	22.6	329	2	Aar51284 L-fucose
9	339.5	21.0	321	2	Aar27118 Fucose de
10	337.5	20.8	306	8	Ad844757 Bacterial
11	329.5	20.4	304	6	Abu17641 Protein e
12	327	20.2	325	8	Adn22222 Bacterial
13	322	19.9	308	8	Adn24982 Bacterial
14	322	19.9	329	6	Abu29907 Protein e
15	317	19.6	367	7	Abu77872 Pseudomon
16	316	19.5	307	8	Ad827427 Bacterial
17	316	19.5	346	6	Abu27806 Protein e
18	314	19.4	325	5	Abp65908 Bifidobac
19	313	19.3	326	6	Aau34593 E. coli c
20	313	19.3	326	6	Abu15320 Protein e
21	313	19.3	326	8	Adn18209 Bacterial
22	311	19.2	346	5	Aam49053 Coriolus
23	310.5	19.2	302	6	Abu43680 Protein e
24	310	19.1	311	6	Abu17621 Protein e
25	307.5	19.0	311	8	Ad830346 Bacterial

26	306.5	18.9	332	6	ABU47024	Abu47024 Protein e
27	305.5	18.9	332	4	AAU36253	Pseudomon
28	305.5	18.9	332	6	ABU47649	Protein e
29	302.5	18.7	341	7	ABO81330	Pseudomon
30	302	18.7	305	6	ABU17897	Protein e
31	301.5	18.6	339	8	ADS41735	Bacterial
32	300.5	18.6	320	8	ADS30364	Bacterial
33	296.5	18.3	315	6	ABU30072	Protein e
34	296.5	18.3	332	7	ADC97264	E. faeciu
35	296	18.3	334	6	ADA34792	Acinetoba
36	295	18.2	344	8	ADS23006	Bacterial
37	294.5	18.2	314	8	ADS41580	Bacterial
38	294.5	18.2	342	8	ADS24931	Bacterial
39	294	18.2	333	6	ABU16921	Protein e
40	291.5	18.0	302	6	ABU42693	Protein e
41	291.5	18.0	314	5	ABP38281	Staphyloc
42	291.5	18.0	314	8	ADS04717	Staphyloc
43	291	18.0	328	6	ABU34174	Protein e
44	291	18.0	348	2	AAU52837	Escherich
45	291	18.0	348	4	AAU34474	E. coli c

ALIGNMENTS

RESULT 1

AAE11998
ID AAE11998 standard; protein; 319 AA.

XX AAE11998;

AC AAE11998;

XX 18-DEC-2001 (first entry)

DT Arabidopsis thaliana L-galactose dehydrogenase (L-galDH) protein.

DE L-galactose dehydrogenase; L-galDH; transgenic organism; ascorbic acid;

XX herbicide; chromosome 4.

OS Arabidopsis thaliana.

XX WO200172974-A2.

XX 04-OCT-2001.

XX 29-MAR-2001; 2001WO-GB001412.

XX 29-MAR-2000; 2000GB-00007651.

XX (ASCO-) ASCORBEX LTD.

XX Smirnoff N, Wheeler G;

XX WPI; 2001-616482/71.

XX N-ESDB; RAD19526.

PT Novel L-galactose dehydrogenase protein and nucleic acid sequence encoding the protein for producing genetically modified plants and microorganisms with enhanced ability to synthesize ascorbic acid.

Claim 1; Page 57-58; 58pp; English.

The present invention relates to an isolated protein having L-galactose dehydrogenase (L-galDH) biological activity. L-galDH nucleic acid is useful for generating transgenic organisms and modified plants with enhanced ability to synthesize ascorbic acid. L-galDH facilitate the production of a plant that has been genetically modified to express a mutated L-galDH protein which is resistant to herbicides that act against the naturally occurring L-galDH and to identify and/or design compounds that are inhibitors of L-galDH. The compounds can be used, for e.g. in a herbicide which acts on L-galDH and damages or kills plants that express the enzyme. The present sequence is Arabidopsis thaliana L-galactose dehydrogenase (L-galDH) protein. The L-galDH gene is located on chromosome 4

```
XX SQ Sequence 319 AA;
Query Match 100.0%; Score 1619; DB 4; Length 319;
Best Local Similarity 100.0%; Pred. No. 2.7e-141;
Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTKIELRALGNTGLKVSAGFGASPLGSGVFGPVAEDDAVATVREAFRLGINFFDTSPPYG 60
DB 1 MTKIELRALGNTGLKVSAGFGASPLGSGVFGPVAEDDAVATVREAFRLGINFFDTSPPYG 60
QY 61 GTLSEKMLGKGLKALQVPRSDYIVATKCRYKEGDFSAERVRKSIDESLERLQLDYVDI 120
DB 61 GTLSEKMLGKGLKALQVPRSDYIVATKCRYKEGDFSAERVRKSIDESLERLQLDYVDI 120
QY 121 LHCHDIEFGSLDQIVSETIPALQKQKGTFRFIGITGLPLDIFTYVLDVRPPGTVDVIL 180
DB 121 LHCHDIEFGSLDQIVSETIPALQKQKGTFRFIGITGLPLDIFTYVLDVRPPGTVDVIL 180
QY 181 SYCHGVNDSTLLDLLPYLKSQGVISASPLAMGLLTEQGPPPEWHSPASPELKSAAV 240
DB 181 SYCHGVNDSTLLDLLPYLKSQGVISASPLAMGLLTEQGPPPEWHSPASPELKSAAV 240
QY 241 AHCKSGKKITKALQYSLANKEISSVLGMSVSVQVEENVAATELSLGMQDQTLSEV 300
DB 241 AHCKSGKKITKALQYSLANKEISSVLGMSVSVQVEENVAATELSLGMQDQTLSEV 300
QY 301 EAILEPVKNLTWPSGIHQN 319
DB 301 EAILEPVKNLTWPSGIHQN 319
RESULT 2
AAU76343
ID AAU76343 standard; protein; 319 AA.
XX AC AAU76343;
XX DT 21-MAY-2002 (first entry)
XX DE A. thaliana L-galactose dehydrogenase (LGDH).
XX KW Ascorbic acid; vitamin C; scurvy; recombinant yeast; enzyme;
XX KW L-galactose dehydrogenase; LGDH; L-galactono-1,4-lactone dehydrogenase;
XX KW AGD; D-arabinose dehydrogenase; ARA; D-arabinono-1,4-lactone oxidase;
XX KW ALO; L-gulono-1,4-lactone oxidase; GLO; aldonolactonase; AL.
XX OS Arabidopsis thaliana.
XX PN WO200210425-A2.
XX PD 07-FEB-2002.
XX PF 02-AUG-2001; 2001WO-GB003485.
XX PR 02-AUG-2000; 2000US-00630983.
XX PA (BIOP-) BIOPOLO SCARL.
XX PA (WHAL/) WHALLEY K.
XX PI Porro D, Sauer M;
XX XX
XX WPÍ; 2002-217125/27.
XX DR N-PSDB; ABK10125.
XX XX
XX PT Generating ascorbic acid or its salt, involves culturing yeast capable of
XX PT converting ascorbic acid precursor into ascorbic acid in medium
XX PT comprising ascorbic acid precursor, and isolating ascorbic acid.
XX XX
XX PS Claim 12; Page 86-87; 95pp; English.
XX XX
XX CC The invention relates to generating ascorbic acid or its salt, involves
XX CC obtaining a recombinant yeast capable of converting an ascorbic acid
```

```
CC precursor into ascorbic acid, culturing the recombinant yeast in a medium
CC comprising an ascorbic acid precursor, thus forming ascorbic acid, and
CC isolating the ascorbic acid. Also include are stabilising ascorbic acid
CC or its salt in a medium, by culturing a yeast in a medium comprising
CC ascorbic acid or its salt and a recombinant yeast functionally
CC transformed with a coding region encoding a protein having an enzyme
CC activity selected from L-galactose dehydrogenase (LGDH), L-galactono-1, 4
CC lactone dehydrogenase (AGD), D-arabinose dehydrogenase (ARA), D-
CC arabinono-1,4-lactone oxidase (ALO), L-gulono-1,4-lactone oxidase (GLO)
CC and aldonolactonase (AL) activity, where the recombinant yeast is capable
CC of converting to ascorbic acid at least about 25% ascorbic acid precursor
CC or is capable of producing at least 20 mg ascorbic acid/L medium, when
CC the yeast is cultured in a medium comprising one ascorbic acid precursor.
CC The ascorbic acid produced (Vitamin C) is a powerful antioxidant, a
CC deficiency of which causes scurvy in humans. The present sequence
XX represents A. thaliana LGDH
SQ Sequence 319 AA;
```

```
Query Match 100.0%; Score 1619; DB 5; Length 319;
Best Local Similarity 100.0%; Pred. No. 2.7e-141;
Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTKIELRALGNTGLKVSAGFGASPLGSGVFGPVAEDDAVATVREAFRLGINFFDTSPPYG 60
DB 1 MTKIELRALGNTGLKVSAGFGASPLGSGVFGPVAEDDAVATVREAFRLGINFFDTSPPYG 60
QY 61 GTLSEKMLGKGLKALQVPRSDYIVATKCRYKEGDFSAERVRKSIDESLERLQLDYVDI 120
DB 61 GTLSEKMLGKGLKALQVPRSDYIVATKCRYKEGDFSAERVRKSIDESLERLQLDYVDI 120
QY 121 LHCHDIEFGSLDQIVSETIPALQKQKGTFRFIGITGLPLDIFTYVLDVRPPGTVDVIL 180
DB 121 LHCHDIEFGSLDQIVSETIPALQKQKGTFRFIGITGLPLDIFTYVLDVRPPGTVDVIL 180
QY 181 SYCHGVNDSTLLDLLPYLKSQGVISASPLAMGLLTEQGPPPEWHSPASPELKSAAV 240
DB 181 SYCHGVNDSTLLDLLPYLKSQGVISASPLAMGLLTEQGPPPEWHSPASPELKSAAV 240
QY 241 AHCKSGKKITKALQYSLANKEISSVLGMSVSVQVEENVAATELSLGMQDQTLSEV 300
DB 241 AHCKSGKKITKALQYSLANKEISSVLGMSVSVQVEENVAATELSLGMQDQTLSEV 300
QY 301 EAILEPVKNLTWPSGIHQN 319
DB 301 EAILEPVKNLTWPSGIHQN 319
```

```
RESULT 3
AAG16520
ID AAG16520 standard; protein; 357 AA.
XX AC AAG16520;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 17197.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX PR 25-FEB-1999; 99US-0121825P.
XX PR 05-MAR-1999; 99US-0123180P.
XX PR 09-MAR-1999; 99US-0123548P.
```

PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131443P.
PR 30-APR-1999; 99US-0132408P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0132866P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135622P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145122P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 22-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.

```
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 97.8%; Score 1583; DB 3; Length 357;
Best Local Similarity 99.7%; Pred. No. 6.8e-138;
Matches 311; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 8 ALGNTGLKVSAGFGASPLGVSFGPVAEDDAVATVREAFRLGINFDTSPYYGGTSLSEKM 67
Db 46 ALGNTGLKVTAVGFGASPLGVSFGPVAEDDAVATVREAFRLGINFDTSPYYGGTSLSEKM 105
QY 68 LGKGLKALQVPSDIYVATKCRYKEGDFDSAEVRKSDSLERLQDLYVDILCHDIE 127
Db 106 LGKGLKALQVPSDIYVATKCRYKEGDFDSAEVRKSDSLERLQDLYVDILCHDIE 165
QY 128 FGSLDQIVSETIPALQKLQKQKTRFIGITGLPLDIFTYVLDVPPGTVDTVLSYCHYGV 187
Db 166 FGSLDQIVSETIPALQKLQKQKTRFIGITGLPLDIFTYVLDVPPGTVDTVLSYCHYGV 225
QY 188 NDSTLLDLLPYLKSQGVVISASPLAMGLLTQGGPPHPASPBLKSASKAVAHCKSKG 247
Db 226 NDSTLLDLLPYLKSQGVVISASPLAMGLLTQGGPPHPASPBLKSASKAVAHCKSKG 285
QY 248 KITTKALQYSLANKEISSVLVGMSSVSOVEENVAATVLESLGMDQETLSEVAILRPV 307
Db 286 KITTKALQYSLANKEISSVLVGMSSVSOVEENVAATVLESLGMDQETLSEVAILRPV 345
QY 308 KNLTPWPSGIHQN 319
Db 346 KNLTPWPSGIHQN 357

RESULT 4
AAG16521
ID AAG16521 standard; protein; 253 AA.
XX
AC AAG16521;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 17198.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; Genetic mapping; gene expression control; promoter;
```

```
KW termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-00301439.
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 28-APR-1999; 99US-0130891P.
XX 30-APR-1999; 99US-0132048P.
XX 04-MAY-1999; 99US-0132407P.
XX 05-MAY-1999; 99US-0132484P.
XX 06-MAY-1999; 99US-0132485P.
XX 06-MAY-1999; 99US-0132486P.
XX 07-MAY-1999; 99US-0132487P.
XX 11-MAY-1999; 99US-0132863P.
XX 14-MAY-1999; 99US-0134218P.
XX 14-MAY-1999; 99US-0134219P.
XX 14-MAY-1999; 99US-0134221P.
XX 18-MAY-1999; 99US-0134370P.
XX 19-MAY-1999; 99US-0134768P.
XX 20-MAY-1999; 99US-0134941P.
XX 21-MAY-1999; 99US-0135124P.
XX 24-MAY-1999; 99US-0135353P.
XX 25-MAY-1999; 99US-0135629P.
XX 27-MAY-1999; 99US-0136021P.
XX 27-MAY-1999; 99US-0136392P.
XX 28-MAY-1999; 99US-0136782P.
XX 01-JUN-1999; 99US-0137222P.
XX 03-JUN-1999; 99US-0137528P.
XX 04-JUN-1999; 99US-0137502P.
XX 07-JUN-1999; 99US-0137724P.
XX 08-JUN-1999; 99US-0138094P.
XX 10-JUN-1999; 99US-0138540P.
XX 10-JUN-1999; 99US-0138847P.
XX 14-JUN-1999; 99US-0139119P.
XX 16-JUN-1999; 99US-0139452P.
XX 16-JUN-1999; 99US-0139453P.
XX 17-JUN-1999; 99US-0139492P.
XX 18-JUN-1999; 99US-0139454P.
XX 18-JUN-1999; 99US-0139455P.
XX 18-JUN-1999; 99US-0139456P.
XX 18-JUN-1999; 99US-0139457P.
XX 18-JUN-1999; 99US-0139458P.
XX 18-JUN-1999; 99US-0139459P.
XX 18-JUN-1999; 99US-0139460P.
XX 18-JUN-1999; 99US-0139461P.
XX 18-JUN-1999; 99US-0139462P.
XX 18-JUN-1999; 99US-0139463P.
XX 18-JUN-1999; 99US-0139750P.
XX 18-JUN-1999; 99US-0139763P.
XX 21-JUN-1999; 99US-0139817P.
XX 22-JUN-1999; 99US-0139899P.
XX 23-JUN-1999; 99US-0140353P.
XX 23-JUN-1999; 99US-0140354P.
XX 24-JUN-1999; 99US-0140695P.
```

PR 28-JUN-1999;	99US-0140823P.
PR 29-JUN-1999;	99US-0140991P.
PR 30-JUN-1999;	99US-0141287P.
PR 01-JUL-1999;	99US-0141842P.
PR 01-JUL-1999;	99US-0142154P.
PR 02-JUL-1999;	99US-0142055P.
PR 06-JUL-1999;	99US-0142300P.
PR 08-JUL-1999;	99US-0142803P.
PR 09-JUL-1999;	99US-0142920P.
PR 12-JUL-1999;	99US-0142977P.
PR 13-JUL-1999;	99US-0143542P.
PR 14-JUL-1999;	99US-0143624P.
PR 15-JUL-1999;	99US-0144005P.
PR 16-JUL-1999;	99US-0144085P.
PR 16-JUL-1999;	99US-0144086P.
PR 19-JUL-1999;	99US-0144325P.
PR 19-JUL-1999;	99US-0144331P.
PR 19-JUL-1999;	99US-0144332P.
PR 19-JUL-1999;	99US-0144333P.
PR 19-JUL-1999;	99US-0144334P.
PR 19-JUL-1999;	99US-0144335P.
PR 20-JUL-1999;	99US-0144352P.
PR 20-JUL-1999;	99US-0144632P.
PR 20-JUL-1999;	99US-0144884P.
PR 21-JUL-1999;	99US-0144814P.
PR 21-JUL-1999;	99US-0145086P.
PR 21-JUL-1999;	99US-0145088P.
PR 22-JUL-1999;	99US-0145085P.
PR 22-JUL-1999;	99US-0145087P.
PR 22-JUL-1999;	99US-0145088P.
PR 22-JUL-1999;	99US-0145089P.
PR 22-JUL-1999;	99US-0145192P.
PR 23-JUL-1999;	99US-0145145P.
PR 23-JUL-1999;	99US-0145218P.
PR 23-JUL-1999;	99US-0145224P.
PR 26-JUL-1999;	99US-0145276P.
PR 27-JUL-1999;	99US-0145913P.
PR 27-JUL-1999;	99US-0145918P.
PR 27-JUL-1999;	99US-0145919P.
PR 28-JUL-1999;	99US-0145951P.
PR 02-AUG-1999;	99US-0146386P.
PR 02-AUG-1999;	99US-0146388P.
PR 02-AUG-1999;	99US-0146389P.
PR 03-AUG-1999;	99US-0147038P.
PR 04-AUG-1999;	99US-0147204P.
PR 04-AUG-1999;	99US-0147302P.
PR 05-AUG-1999;	99US-0147192P.
PR 05-AUG-1999;	99US-0147260P.
PR 06-AUG-1999;	99US-0147303P.
PR 09-AUG-1999;	99US-0147493P.
PR 09-AUG-1999;	99US-0147935P.
PR 10-AUG-1999;	99US-0148171P.
PR 11-AUG-1999;	99US-0148319P.
PR 12-AUG-1999;	99US-0148344P.
PR 13-AUG-1999;	99US-0148565P.
PR 13-AUG-1999;	99US-0148684P.
PR 16-AUG-1999;	99US-0149368P.
PR 17-AUG-1999;	99US-0149175P.
PR 18-AUG-1999;	99US-0149425P.
PR 20-AUG-1999;	99US-0149722P.
PR 20-AUG-1999;	99US-0149723P.
PR 20-AUG-1999;	99US-0149929P.
PR 23-AUG-1999;	99US-0149902P.
PR 23-AUG-1999;	99US-0149930P.
PR 25-AUG-1999;	99US-0150566P.
PR 26-AUG-1999;	99US-0150884P.
PR 27-AUG-1999;	99US-0151065P.
PR 27-AUG-1999;	99US-0151066P.
PR 27-AUG-1999;	99US-0151080P.
PR 30-AUG-1999;	99US-0151303P.
PR 31-AUG-1999;	99US-0151438P.
PR 01-SEP-1999;	99US-0151930P.
PR 07-SEP-1999;	99US-0152363P.
PR 10-SEP-1999;	99US-0153070P.
PR 13-SEP-1999;	99US-0153758P.
PR 15-SEP-1999;	99US-0154018P.
PR 16-SEP-1999;	99US-0154039P.
PR 20-SEP-1999;	99US-0154779P.
PR 22-SEP-1999;	99US-0155139P.
PR 23-SEP-1999;	99US-0155486P.
PR 24-SEP-1999;	99US-0155659P.
PR 28-SEP-1999;	99US-0156458P.
PR 29-SEP-1999;	99US-0156596P.
PR 04-OCT-1999;	99US-0157117P.
PR 05-OCT-1999;	99US-0157753P.
PR 06-OCT-1999;	99US-0157865P.
PR 07-OCT-1999;	99US-0158029P.
PR 08-OCT-1999;	99US-0158232P.
PR 12-OCT-1999;	99US-0158369P.
PR 13-OCT-1999;	99US-0159293P.
PR 13-OCT-1999;	99US-0159294P.
PR 13-OCT-1999;	99US-0159295P.
PR 14-OCT-1999;	99US-0159329P.
PR 14-OCT-1999;	99US-0159330P.
PR 14-OCT-1999;	99US-0159331P.
PR 14-OCT-1999;	99US-0159637P.
PR 18-OCT-1999;	99US-0159638P.
PR 18-OCT-1999;	99US-0159584P.
PR 21-OCT-1999;	99US-0160741P.
PR 21-OCT-1999;	99US-0160767P.
PR 21-OCT-1999;	99US-0160768P.
PR 21-OCT-1999;	99US-0160770P.
PR 21-OCT-1999;	99US-0160814P.
PR 21-OCT-1999;	99US-0160815P.
PR 22-OCT-1999;	99US-0160980P.
PR 22-OCT-1999;	99US-0160981P.
PR 22-OCT-1999;	99US-0160989P.
PR 25-OCT-1999;	99US-0161404P.
PR 25-OCT-1999;	99US-0161405P.
PR 25-OCT-1999;	99US-0161406P.
PR 26-OCT-1999;	99US-0161359P.
PR 26-OCT-1999;	99US-0161360P.
PR 26-OCT-1999;	99US-0161361P.
PR 28-OCT-1999;	99US-0161920P.
PR 28-OCT-1999;	99US-0161922P.
PR 28-OCT-1999;	99US-0161993P.
PR 29-OCT-1999;	99US-0162142P.
Query Match 79.4%; Score 1286; DB 3; Length 253;	
Best Local Similarity 100.0%; Pred. No. 1.4e-110;	
Matches 253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 67	MLGKGLKALQVPRSDYIVATKCGRYKEGDFPSAERVRKSIDESLERLQLDYVDILHCHDI 126
DB 1	MLGKGLKALQVPRSDYIVATKCGRYKEGDFPSAERVRKSIDESLERLQLDYVDILHCHDI 60
QY 127	BFGLDQIVSETIPALOKLQKQEGKTRFITGLDIFTYYLDRVPPGTVDVILSYCHYG 186
DB 61	BFGLDQIVSETIPALOKLQKQEGKTRFITGLDIFTYYLDRVPPGTVDVILSYCHYG 120
QY 187	VNDSTLLDLLPYLKS KGVVISASPLAMGLLTEQOPPEWHPASPELKSASKAANAHCXSK 246
DB 121	VNDSTLLDLLPYLKS KGVVISASPLAMGLLTEQOPPEWHPASPELKSASKAANAHCXSK 180
QY 247	GKKTTLALQYSLANKKEISSVLVGMSSVSQVEENVAATELESIGMDQETLSEVEAILLEP 306
DB 181	GKKTTLALQYSLANKKEISSVLVGMSSVSQVEENVAATELESIGMDQETLSEVEAILLEP 240
QY 307	VKNLTWPSGIHQN 319
DB 241	VKNLTWPSGIHQN 253
RESULT 5	
ABB67077	
ID ABB67077 standard; protein; 345 AA.	

XX AC ABB67077;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 28023.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL11180.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions.
XX PS Disclosure; SEQ ID NO 28023; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX CC
XX CC Sequence 345 AA;
XX CC
Query Match 36.7%; Score 593.5; DB 4; Length 345;
Best Local Similarity 42.5%; Pred. No. 3.6e-46; Mismatches 99; Indels 11; Gaps 7;
Matches 128; Conservative 63;
QY 3 KIELRALGNTGLKVSAGFGASPLGVSFGPVAEDDAVATVREAFRLGINFDTSPYGGT 62
Db 21 RMEYRNLTGTLQVSKVSFGGALCANYGFDLE-EGIKTVHEAVKSGINYIDTAPYGGG 79
QY 63 LSEKMLGKGLKALQVPRSDYIVATKGRYK----EGFDPFAERVKSIDESLERLQLDYV 118
Db 80 RSEEVGLGALK--DVPRESYIATKVARYELDYDMFDFSAKKTRESVEKSLKLGLDYV 137
QY 119 DILHCHDIEFG-SLDQIVSETIPALQKQKQKTRFGITGLPLDIFTYVLDVPPGTV 177
Db 138 DVIQIHDFBAKDLDIVINETLPTLEQLVKEGKARFIVGSAYPSVLKEFLRT-AGRLD 196
QY 178 VILSYCHYGVNDSTLIDLPLYLKSGVGVISASPLAMGLLTQGGPPEWHPASPPLKSK 237
Db 197 TVLTARYTLTDETLLEVLDFKSNLGVICAAHALGLLTWAGQPWHPASDEQKAIAR 256
QY 238 AAVACKSGKKITKIALQYSIAN-KETSSVLVGMSSVSQVEENVAATVEESLGMDOET 296
Db 257 KASEVCKERGVLGLKAMTYTMSGLPEVSTFTLGMQTRQLLRINLDA-NEVGLSKDEQEV 315
QY 297 L 297

Db 316 L 316
RESULT 6
ABB71229
ID ABB71229 standard; protein; 342 AA.
XX AC ABB71229;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 40479.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL15332.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions.
XX PS Disclosure; SEQ ID NO 40479; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX CC
XX CC Sequence 342 AA;
XX CC
Query Match 33.6%; Score 544.5; DB 4; Length 342;
Best Local Similarity 36.6%; Pred. No. 1.2e-41;
Matches 117; Conservative 71; Mismatches 117; Indels 15; Gaps 7;
QY 3 KIELRALGNTGLKVSAGFGASPLGVSFGPVAEDDAVATVREAFRLGINFDTSPYGG 61
Db 21 RMEYRNLTGTLQVSKVSFGGALCANYGFDLE-EGIKTVHEAVKSGINYIDTAPYGG 80
QY 62 TLSEKMLGKGLKALQVPRSDYIVATKGRYK----EGFDPFAERVKSIDESLERLQLDY 117
Db 81 GKSEELLGQALK--DVPREAYIATKVARYELDPNNMFDYTAARAKRESVKSLKLGLD 138
QY 118 VDILHCHDIEFG-SLDQIVSETIPALQKQKQKTRFGITGLPLDIFTYVLDVPPGTV 176
Db 139 VDLVQLVHDVDAAPSLDMLVNLNITIPVLEFYVQAGKARFIVTAYDVVLKECAER-GKGRI 197
QY 177 DVILSYCHYGVNDSTLIDLPLYLKSGVGVISASPLAMGLLTQGGPPEWHPASPPLKSK 236
Db 198 QVLTNARYTLTDLNLTLLRHMKAFQBMGVGVVCAAAHSLGLLSNAGPQSWHPGSPPELLAVG 257


```

Db      111 DILQVHDVDAAPNLDIVLNETIPVLVEYVQAGKARFIGVTAYDVDVLEKAER-CKGRIQ 167
Qy      178 VILSYCHGVNDSTLLDLLPYLKSGVGVISASPLAMGLLTEQGPEWHPASPELSASK 237
Db      170 VWLNARYTLLDNTLLRYMKDFQKWGGVVCAAHAHSLGLLRNAGPHASHPGSQEITLAVAK 229
Qy      238 AAVAHCSKKKITKLALQVSLANKEISSVLVGHSSVSQVENVAANTVELBSLGM---DO 294
Db      230 RGAEICCOORVELGKLAMYTYTMQLDGAATFLIGIPNRKCLRINLDAIFD----GLTSHEG 285
Qy      295 ETL 297
Db      286 EVL 288

RESULT 8
AARS51284
ID   AARS51284 standard; protein, 329 AA.
AC   AARS51284;
XX
XX
DT   16-OCT-2003 (revised)
DT   14-DEC-1994 (first entry)
XX
XX   L-fucose dehydrogenase.
DE
XX
XX   L-fucose dehydrogenase; L-FDH; enzyme.
XX
XX   Pseudomonas sp; No.1143.
OS
XX
XX   JF06090765-A.
XX
XX   05-APR-1994.
XX
XX   11-SEP-1992;    92JP-00243372.
PF
XX
XX   11-SEP-1992;    92JP-00243372.
PR
XX
XX   (KIKK ) KIKKOMAN CORP.
XX   (NODA ) ZH NODA SANGYO KAGAKU KENKYUSHO.
PA
XX
XX   WPI; 1994-146990/18.
DR
XX
XX   N-PSDB; AAQ62089.
XX
XX
XX   Novel L-fucose dehydrogenase gene - a new recombinant DNA and the prepn.
XX   of L-FDH using a L-FDH vector.
PT
XX
XX   Claim 1; Page 5; 6pp; Japanese.
PS
XX
XX   The L-FDH gene is new and can be used for the recombinant prodn. of the
CC   enzyme. (Updated on 16-OCT-2003 to standardise OS field)
CC
XX
XX   Sequence 329 AA;
SQ

Query Match          22.6%; Score 366; DB 2; Length 329;
Best Local Similarity 30.7%; Pred.No.4.2e-25;
Matches 103; Conservative 62; Mismatches 122; Indels 48; Gaps 9

Qy      8 ALGNWTKLVSAVGESGSLGSVFGEVADDAVATVREAFLRGINFDDTPSYGGTLLSEKM 67
Db      7 AAAAGAIAI PALGYGANVGNLFALSDDEAWALEAAWDAGIRYYIDTAPHYGLGLESKR 66
Qy      68 LGKGLKALQVRSYDIYVATKGR-----YKEGFDFSAEVR 103
Db      67 LGAFLOT--KRFDEFVSTKAGRLLRNPERRPSGLDTDNDHFVPDDLREWDFTTEQGI 124
Qy      104 KSIDESLERIQDYVDILHCHDIEFGSLDQIVSETIPALQKLKGEGTRFIGITGLPLDI 163
Db      125 ASIAESQERLGDRIDLILYHDPERHDLDALASAFPALAEKVRAEGVVKAIIGISMVSDA 184
Qy      164 FTVVLDVRVPGTVVILSYCHGVNDSTLL-----DLLPYLKSGVGVISASPLAMGLL 217
Db      185 LTIRA---VREADLILMWAGRY-----TLLEQPAATEVLPAEAENATGIVASVFNSSLG 236

```

```
QY 218 TQGGP-----EWHPASPELKSASKAANAHAHCKSKGKTKLQVSLANKEISSVLVQMS 272
D 237 AQSEPKRDRGEYQGLPDELMDRLVRIIAICRNHDPVLPAAAIQPLQSAIVRSVVVGG 296
QY 273 SVSQVEENVA-AVTELESIGMDQETLSEVEATLEP 306
D 297 RPAQLTONAYAALEIPA-GUWAE-LAEARLIPTP 329

RESULT 9
AAR27118
ID AAR27118 standard; protein; 321 AA.
XX
AC AAR27118;
XX
DT 24-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 01-MAR-1993 (first entry)
XX
DE Fucose dehydrogenase.
XX
KW Arthrobacter oxidans; F1; induction; assay.
XX
OS Arthrobacter oxydans; F1.
XX
PN BP506262-A1.
XX
PD 30-SEP-1992.
XX
PF 13-MAR-1992; 92EP-00302170.
XX
PR 29-MAR-1991; 91JP-00089184.
XX
PA (TAKI ) TAKARA SHUZO CO LTD.
XX
PI Mitta M, Kotani H, Sakai T, Kato I;
XX
WPI; 1992-325548/40.
DR N-PSDB; AAQ28895.
XX
PT Isolated gene encoding L-fucose dehydrogenase - useful for prodn. of
PT enzyme by genetic engineering.
XX
PS Disclosure; Page 8; 16pp; English.
XX
CC Genomic DNA from Arthrobacter oxydans F1 was subjected to restriction
CC enzyme analysis and the N-terminal amino acid sequence of L-fucose
CC dehydrogenase determined. A degenerate probe was synthesised based on
CC this amino acid sequence. The probe was used to screen an Arthrobacter
CC cDNA library to isolate a L-fucose dehydro- genase clone. The isolation
CC of such a clone provides a convenient method for prodn. of L-fucose
CC dehydrogenase without the need for induction by L-fucose. The probe may
CC be used to evaluate the extent of expression of L-fucose dehydrogenase.
CC The DNA sequence is widely used to assay L-fucose levels. The protein
CC sequence of L-fucose dehydrogenase may be used to produce antibodies.
CC (Updated on 25-MAR-2003 to correct PN field.) (Updated on 24-OCT-2003 to
CC standardise OS field)
XX
SQ Sequence 321 AA;

Query Match 21.0%; Score 339.5; DB 2; Length 321;
Best Local Similarity 30.9%; Pred. No. 1.2e-22;
Matches 99; Conservative 58; Mismatches 110; Indels 53; Gaps 11;

QY 12 TGLKYSAGFGASPLGSGVGPVADDAVATVREAFRLGINFDTSPYGGTLSERMLGKG 71
D 9 TDVLLPVLGFGAGPAGNLVGPVPEQEAASAAWEGGVRYFTARHGLGUSERMGA 68
QY 72 LKALQVPRGDIYATKGRY-----KEGF-----DFAERVKSIDESL 110
D 69 LAGQQ--RDSYVLSTKIGRLLRPNPSPQCKDTEGDFVDPDLIRVDRYSDRGVLRISL 126

111 ERLQIDYVDILHCHDIEFGSLDQIVSETIPALQKLKQEGKTRFIGI-----TGLPLDIPTY 166
D 127 QRLGTDRIDIVYIHDPD--DYWTEAVEGAAPALSALRDEGVIRAWGAGMNSGM-LHREIT 184
QY 167 VLDVRPPGTVDVILSYCHYGVNDSTLL-----DLLPYLKSQGVGVISASPLAMGLLITEQ 220
D 185 ETD-----IDVMLAGRY-----TLLEQGAQDILLPACLERGVGVNVGVFNSGLLSKE 233
QY 221 GPP-----EWHPASPELKSASKAANAHAHCKSKGKTKLQVSLANKEISSVLVQMSVS 275
D 234 RPAANATVYAPAQEILLDRANILLADVCESHGTTLPAAALHFPYQHFAVTSVILGMRTPA 293

276 QVEEN--VAAVTELESIGMD 293
D 294 QVKQNLDIASQTVPDQLWAD 313

RESULT 10
ADS44757
ID ADS44757 standard; protein; 306 AA.
XX
AC ADS44757;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polypeptide #23187.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
WPI; 2004-061375/06.
XX
New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
Claim 1; SEQ ID NO 23187; 122pp; English.
XX
The invention relates to a recombinant DNA construct comprising a
promoter functional in a plant cell, where the promoter is positioned to
provide for expression of a polynucleotide encoding a polypeptide from a
microbial source. The invention also relates to a transformed plant
comprising the recombinant DNA construct and a method of producing a
transformed plant having an improved property. The plant is a crop plant
such as maize or soybean. The method of producing a transformed plant
having an improved property comprises transforming a plant with the
recombinant DNA construct and growing the transformed plant, where the
polynucleotide or polypeptide is useful for improving plant properties.
The recombinant DNA construct is useful for producing plants with
improved plant properties, e.g. improved cold, heat or drought tolerance,
```


RESULT 12	
ADN22222	
ID	ADN22222 standard; protein; 325 AA.
XX	
XX	ADN22222;
XX	
DT	02-DEC-2004 (first entry)
XX	
DE	Bacterial polypeptide #4875.
XX	
KW	Recombinant DNA construct; transformed plant; improved plant property;
KW	cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW	pathogen tolerance; pest tolerance; plant disease resistance;
KW	cell cycle pathway modification; plant growth regulator;
KW	homologous recombination; seed oil yield; protein yield; carbohydrate;
KW	nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW	bacterial polypeptide.
XX	
OS	Bacteria.
XX	
XX	US2003233675-A1.
PN	
XX	18-DEC-2003.
XX	
PF	20-FEB-2003; 2003US-00369493.
XX	
PR	21-FEB-2002; 2002US-0360039P.
XX	
PA	(CAOY/) CAO Y.
PA	(HINK/) HINKLE G J.
PA	(SLAT/) SLATER S C.
PA	(CHEN/) CHEN X.
PA	(GOLD/) GOLDMAN B S.
XX	
PI	Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX	
DR	WPI; 2004-061375/06.
XX	
PT	New recombinant DNA construct comprising a promoter positioned to provide
PT	for expression of a polynucleotide encoding a polypeptide from a
PT	microbial source, useful for producing plants with improved properties.
XX	
PS	Claim 1; SEQ ID NO 4875; 122pp; English.
XX	
CC	The invention relates to a recombinant DNA construct comprising a
CC	promoter functional in a plant cell, where the promoter is positioned to
CC	provide for expression of a polynucleotide encoding a polypeptide from a
CC	microbial source. The invention also relates to a transformed plant
CC	comprising the recombinant DNA construct and a method of producing a
CC	transformed plant having an improved property. The plant is a crop plant
CC	such as maize or soybean. The method of producing a transformed plant
CC	having an improved property comprises transforming a plant with the
CC	recombinant DNA construct and growing the transformed plant, where the
CC	polynucleotide or polypeptide is useful for improving plant properties.
CC	The recombinant DNA construct is useful for producing plants with
CC	improved plant properties, e.g. improved cold, heat or drought tolerance,
CC	tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC	increased resistance to plant disease, better growth rate by modification
CC	of the cell cycle pathway with plant growth regulators, increased rate of
CC	homologous recombination, modified seed oil or protein yield and/or
CC	content, improved yield by modification of carbohydrate, nitrogen or
CC	phosphorus use and/or uptake, by modification of photosynthesis or by
CC	providing improved plant growth and development under at least one stress
CC	condition, improved lignin production or improved galactomannan
CC	production. This sequence represents a bacterial polypeptide used in the
CC	scope of the invention. Note: The sequence data for this patent did not
CC	form part of the printed specification but was obtained in electronic
CC	format from USPTO at seqdata.uspto.gov/sequence.html .
XX	
SQ	Sequence 325 AA;

[illegible]

RESULT 13	
ADN24982	
ID	ADN24982 standard; protein; 308 AA.
XX	
XX	
AC	ADN24982;
XX	
XX	
02-DEC-2004	(first entry)
DT	
XX	
DE	Bacterial polypeptide #7635.
XX	
XX	Recombinant DNA construct; transformed plant; improved plant property;
KW	cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW	pathogen tolerance; pest tolerance; plant disease resistance;
KW	cell cycle pathway modification; plant growth regulator;
KW	homologous recombination; seed oil yield; protein yield; carbohydrate;
KW	nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW	bacterial polypeptide.
XX	
XX	
OS	Bacteria.
XX	
PN	US2003233675-A1.
XX	
PD	18-DEC-2003.
XX	
XX	
PF	20-FEB-2003; 2003US-00369493.
XX	
PR	21-FEB-2002; 2002US-0360039P.
XX	
PA	(CAOY/) CAO Y.
PA	(HINK/) HINKLE G J.
PA	(SLAT/) SLATER S C.
PA	(CHEN/) CHEN X.
PA	(GOLD/) GOLDMAN B S.
XX	
XX	
FI	Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX	
DR	WPI; 2004-061375/06.
XX	
XX	
PT	New recombinant DNA construct comprising a promoter positioned to provide
PT	for expression of a polynucleotide encoding a polypeptide from a
PT	microbial source, useful for producing plants with improved properties.
XX	
XX	Claim 1; SEQ ID NO 7635; 122pp; English.
PS	
XX	
CC	The invention relates to a recombinant DNA construct comprising a

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2005, 08:32:13 ; Search time 28 Seconds
(without alignments)
850.466 Million cell updates/sec

Title: US-10-606-300-11

Perfect score: 1619

Sequence: 1 MTKIELRALGNTGLKVSAG.....VEAILEPVKNLTWPSGIHQN 319

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_AA*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1619	100.0	319	US-09-630-983A-11	Sequence 11, Appl
2	317	19.6	367	US-09-252-991A-26618	Sequence 26618, A
3	314.5	19.4	347	US-09-902-540-10503	Sequence 10503, A
4	302.5	18.7	341	US-09-252-991A-30076	Sequence 30076, A
5	300.5	18.6	332	US-09-902-540-10805	Sequence 10805, A
6	296.5	18.3	332	US-09-107-532A-6891	Sequence 6891, Ap
7	296	18.3	334	US-09-328-352-6079	Sequence 6079, Ap
8	291.5	18.0	314	US-09-134-001C-3126	Sequence 3126, Ap
9	291	18.0	348	US-09-673-198-4	Sequence 4, Appl
10	277.5	17.1	364	US-09-489-039A-12541	Sequence 12541, A
11	277	17.1	334	US-09-724-623-86	Sequence 86, Appl
12	274	16.9	350	US-09-489-039A-9552	Sequence 9552, Ap
13	271	16.7	328	US-09-489-039A-12803	Sequence 12803, A
14	264	16.3	358	US-09-252-991A-31535	Sequence 31535, A
15	254	15.7	347	US-09-489-039A-9489	Sequence 9489, Ap
16	252.5	15.6	325	US-09-902-540-12689	Sequence 12689, A
17	250	15.4	367	US-09-949-016-6399	Sequence 6399, Ap
18	249	15.4	324	US-09-949-016-10971	Sequence 10971, A
19	247	15.3	353	US-09-328-352-6721	Sequence 6721, Ap
20	245.5	15.2	314	US-09-134-001C-5405	Sequence 5405, Ap
21	245	15.1	329	US-08-606-143-2	Sequence 2, Appl
22	244.5	15.1	351	US-09-252-991A-19370	Sequence 19370, A
23	242	14.9	329	US-08-606-143-1	Sequence 1, Appl
24	242	14.9	329	US-08-606-143-3	Sequence 3, Appl
25	234.5	14.5	319	US-09-902-540-11071	Sequence 11071, A
26	233	14.4	314	US-09-902-540-14628	Sequence 14628, A
27	221	13.7	330	US-09-679-279-10	Sequence 10, Appl

28 219.5 13.6 290 4 US-09-602-787A-670 Sequence 670, App
29 219.5 13.6 380 4 US-09-248-796A-17336 Sequence 17336, A
30 218.5 13.5 316 4 US-09-134-000C-5598 Sequence 5598, Ap
31 218.5 13.5 317 4 US-09-252-991A-22776 Sequence 22776, A
32 218 13.5 357 4 US-09-252-991A-24365 Sequence 24365, A
33 217.5 13.4 353 4 US-10-092-263-8 Sequence 8, Appl
34 217 13.4 1114 2 US-08-576-626A-31 Sequence 31, Appl
35 216 13.3 342 4 US-09-949-016-7654 Sequence 7654, Ap
36 214.5 13.2 310 4 US-09-252-991A-32639 Sequence 32639, A
37 214 13.2 306 4 US-09-902-540-9859 Sequence 9859, Ap
38 214 13.2 342 4 US-09-538-092-776 Sequence 776, App
39 213.5 13.2 353 4 US-09-248-796A-17342 Sequence 17342, A
40 211 13.0 398 4 US-09-489-039A-9541 Sequence 9541, Ap
41 209 12.9 330 4 US-09-724-797-28 Sequence 28, Appl
42 208.5 12.9 291 4 US-09-107-532A-5063 Sequence 5063, Ap
43 208.5 12.4 255 4 US-09-583-110-3418 Sequence 3418, Ap
44 200.5 12.4 321 4 US-09-489-039A-13361 Sequence 13361, A
45 200 12.4 366 4 US-09-248-796A-17341 Sequence 17341, A

ALIGNMENTS

RESULT 1

US-09-630-983A-11

; Sequence 11, Application US/09630983A

; Patent No. 6630330

; GENERAL INFORMATION:

; APPLICANT: Porro, Danilo

; APPLICANT: Sauer, Michael

; TITLE OF INVENTION: Ascorbic Acid Production from Yeast

; FILE REFERENCE: 2028.594000

; CURRENT APPLICATION NUMBER: US/09/630,983A

; CURRENT FILING DATE: 2000-08-02

; NUMBER OF SEQ ID NOS: 26

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 11

; LENGTH: 319

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-630-983A-11

Query Match 100.0%; Score 1619; DB 4; Length 319;

Best Local Similarity 100.0%; Pred. No. 5.5e-162;

Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MTKIELRALGNTGLKVSAGVFGASPLGSGVFGPVAEDDAVATVREAFRLGNEFFDTSPPYG	60
Db	1	MTKIELRALGNTGLKVSAGVFGASPLGSGVFGPVAEDDAVATVREAFRLGNEFFDTSPPYG	60
Qy	61	GTLSEKMLGKGLKALQVPRSDYIVATKCGRYKEGDFSAERVRKSIDESLERLQLDYVDI	120
Db	61	GTLSEKMLGKGLKALQVPRSDYIVATKCGRYKEGDFSAERVRKSIDESLERLQLDYVDI	120
Qy	121	LHCHDIEFGSLDQIVSETIPALQKLQKQKTRFRTGITGLDIDIFTYVLDVRVPPGVTVIL	180
Db	121	LHCHDIEFGSLDQIVSETIPALQKLQKQKTRFRTGITGLDIDIFTYVLDVRVPPGVTVIL	180
Qy	181	SYCHGVNDSTLLDLLPYLKSGVGVISASPLAMGLLTFEQGPPEWHFASPELKSAAV	240
Db	181	SYCHGVNDSTLLDLLPYLKSGVGVISASPLAMGLLTFEQGPPEWHFASPELKSAAV	240
Qy	241	AHCKSGKKITKALQVSLANKETSSVLVGMSSVQVEENVAATVTELSLGNQDTLSV	300
Db	241	AHCKSGKKITKALQVSLANKETSSVLVGMSSVQVEENVAATVTELSLGNQDTLSV	300
Qy	301	EALIEPVKNLTWPSGIHQN	319
Db	301	EALIEPVKNLTWPSGIHQN	319

RESULT 2

US-09-252-991A-26618

; Sequence 26618, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26618
; LENGTH: 367
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26618

Query Match 19.6%; Score 317; DB 4; Length 367;
Best Local Similarity 29.3%; Pred. No. 1.3e-24;
Matches 84; Conservative 60; Mismatches 103; Indels 40; Gaps 8;

QY 19 VFGFASPLGSGVPGVAEDDAVATVREAFRLGNGINFFDTSPYGGTISEKMLGKGLKALQVP 78
DB 52 IGLGAPLGNMFHPLSEETADATLNAWDAGFRYYDVSPHYGAGLAEQRFRLLSG--KP 109
QY 79 RSDYIVATKCGRY-----KEGDFSAERVKSIDESLERLQDYVD 119
DB 110 RDEVVLSYKVGRLQAPASOPENAKFFVDELPNKRVDPYSDAGARSIEDSLERMGVDRLD 169
QY 120 ILHCHDI---EFGS-----LDQIVSETIPALQKQEGKTRFIGITGLPLDIFTYVLDRV 171
DB 170 VVFIHDVSEDQWPQWREYFOAMNGAAKALTQLRDEGVIRGWLGVNLVPCRLALRQS 229
QY 172 PPGTVDVILSYCHYGV--NDSTLLDLLPYLKSXGKGVISASPLAMGLITEQGPPEWHAS 229
DB 230 DP---NVFLLAGRYSLLEHDEALDTLFTTCQARDVGVVVGGPFNSGVLAGGDHYEQIP 286
QY 230 PEL---KSAKAHVAHCKSKGKTKLALQYSLANKKEISSVLVGMSS 273
DB 287 PQVAQRRLKAAAEHC---GVDLRAAHLFLCLANPVVASVIPGTAN 330

RESULT 3
US-09-902-540-10503
; Sequence 10503, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 10503
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-10503

Query Match 19.4%; Score 314.5; DB 4; Length 347;
Best Local Similarity 26.9%; Pred. No. 2.2e-24;
Matches 87; Conservative 62; Mismatches 119; Indels 55; Gaps 9;

QY 4 IETRALGNTGLKVASVGFASPLGSGVFGPVAEDDAVATVREAFRLGNGINFFDTSPYGGTIL 63

DB 12 VPTNAPGNKFRPTTRIGLGGVAINGPASTSDAQQAQATLEAANAAGVRYFDTSPWYGLCL 71
QY 64 SEKMLGKGLKALQVPRSDYIVATKGR-----YKGEFDFSAERVRK 104
DB 72 SERAFGRFLHTRDA--NEVILSTKVRILTAADAPPKLLMSDASPFNYQYDYSAGVRR 129
QY 105 SIDESLERLQDYVDILHCHDI---EFGS-----LDQIVSETIPALQKQEGKTRFI 154
DB 130 SVEDSLQRLGVSRIDIVFIHDLSPDNQMDKWTYFQAIKGAIPELTKREELIKAW 189
QY 155 GITGLPLDIFTYVLDRVPP-----GTVDVILSYCHYGV--NDSTLLDLLPYLKSXGKGV 205
DB 190 G-----FGVNRAEPALRALEVPDIFLIATQYSLTNHEALSKTFPALEKRGIS 239
QY 206 VISASPLAMGLITEQGPPEWH---PASPELKSASAAVAHCKSKGKTKLALQYSLANK 262
DB 240 VVVGSPLDAGYLACGRNRYLDGTTIPPGVQEKRAARMSAIA--ACHGIDRLTAALQFAAAPS 297
QY 263 EISSVLVGMSSVSQVEENVAAT 285
DB 298 VVSSVIPGARTPEQYKANVQSM 320

RESULT 4

US-09-252-991A-30076
; Sequence 30076, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30076
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30076

Query Match 18.7%; Score 302.5; DB 4; Length 341;
Best Local Similarity 28.8%; Pred. No. 3.8e-23;
Matches 94; Conservative 58; Mismatches 127; Indels 47; Gaps 11;
QY 9 LGNTGLKVASVGFASPLGSGVFGP--VAEDDAVATVREAFRLGNGINFFDTSPYGGTISE 65
DB 24 LGNSGLKVSRLCLGCMYTGDPAPRFPWLDREARPFIREALEAGIDFFDSADIVSTGESE 83
QY 66 KMLGKGLKALQVPRSDYIVATK-----CGRYKEGDFSAERVKSIDESLERLQ 114
DB 84 RILGRAURDF-AQREDLVIAATKAFPMDSRDNAC-----LSRKHLASVDASLRLUG 135
QY 115 LDYVDILHCHDIERFSGSLDQIVSETIPALQKQEGKTRFIGITGLPLDIFTYVLD-RVPP 173
DB 136 TDYLDLFIHTRFD---PETPIETCETLDSLVRAGKRVYLGASSMPANRPMKLAQFQRHH 192
QY 174 GTVDVILSYCHYGV--NDSTLLDLLPYLKSXGKGVISASPLAMGLI-----T 218
DB 193 GLAQFISMOSQYNLIVREDEE--DLVPLCREEGIALTPWSPLARGLLAGARSAGTLRTRT 250
QY 219 EQGPPEWHPASPELKSASAAVAHCKSKGKTKLALQYSLANKKEISSVLVGMSSVSQVE 278
DB 251 DEQAPRWYGGREEVSTGLALEKLAARGLPPAQLALAWLLGRNGVAAPIVGLSRPHHLE 310
QY 279 ENVAATVEISLGMDDQETLSEVEAIL 304
DB 311 DALAALT---LDLAEEECATLEAPL 332


```
RESULT 5
US-09-902-540-10805
; Sequence 10805, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiseland, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 10805
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-10805

Query Match      18.6%; Score 300.5; DB 4; Length 332;
Best Local Similarity 29.0%; Pred. No. 6e-23;
Matches 97; Conservative 59; Mismatches 128; Indels 51; Gaps 11;

Qy      4  IELRALGNTGLKVSAGVGA-SPLGVSFPGVAEDDAVATVREAFRLGINFDTSPYGGT 62
Db      1  MHFRKLGSLVSEISGNITHGS-----QVEEAAACVRAALDVGITTFDTADVAAT 57

Qy      63  LSEKMLGKGLKALQVPRSDYIVATK-----CGRYKEGDFPSAERVKRSIDESLERLQLDYV 118
Db      58  RAEVIGRALKGER--RAGYELFTKYVWPTGPKNDRLGSRKHILESIDGSLRLQTDYV 115

Qy      119  DILHCHDIEFGLSDQIVSETTALQKQKQKTRFIGHTGLPLDIFTYVLDVPPGT--- 175
Db      116  DLYQAHRED---VETPLEETMLAFADIVROGKALYIGVS-----EWTADQIRQGAALA 165

Qy      176  ---VDVTLSYCHYGVNDSTL-LDLLPYLKSQGVVISASPLAMGLLT-----EQGPPW 225
Db      166  RELRVFFISNQPOYSMLRVTEPQVIPASDEAGLQIVWSPIAQGLVTKGLPQAPPAG 225

Qy      226  HPA-----SPELKSASAAVAHCKSKGKITKLALQYSLANKEISSVLVGM 271
Db      236  SRATEANAVRYGITREFTDDVLTTRVQQLVPLAKDVGLSMAQLAVAWVLQNPVSVAIGA 285

Qy      272  SSVSQVEENVAATELSIGMDQETLSEVBAILEP 306
Db      286  SRPEQVHDNVKAA-----GVKLEPELLRRIDAVLGP 316

RESULT 6
US-09-107-532A-6891
; Sequence 6891, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6891:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...332
; SEQUENCE DESCRIPTION: SEQ ID NO: 6891:
US-09-107-532A-6891

Query Match      18.3%; Score 296.5; DB 4; Length 332;
Best Local Similarity 27.5%; Pred. No. 1.6e-22;
Matches 88; Conservative 63; Mismatches 132; Indels 37; Gaps 10;

Qy      5  ELRALGNTGLKVSAGVGA-SPLG--SVFPGVAEDDAVATVREAFRLGINFDTSPYGGT 62
Db      20  EIR-IGHSQVTAEOQLGAGANAVGHNLFDGLEDSTGKQVVRTALNSGINLIDTAYANG 78

Qy      63  LSEKMLGKGLKALQVPRSDYIVATKCGRYK---EGFDFSAERVKRSIDESLERLQLDYVD 119
Db      79  RSEELIGVLKEKEYDSRVVIATKAAHVPNKGRFTDNPSEFLKQSVEDALKRLQTDYID 138

Qy      120  ILHCHDIEFGLSDQIVSETTALQKQKQKTRFIGHTGLPLDIFTYVLDVPPGTVDVI 179
Db      139  IFYTH---FPDESTPKNESVATLHELKEAGKIRAVGVSNFTLE---QLKEANADGYVDV 192

Qy      180  ---LSYCHYGVNDSTLDDLPLKSKGVVISASPLAMGLLT-----EQGPPWHP 227
Db      193  EDKYSLIHQAEK---ELFPYLEKKNKISFVFPPLASGLLTGKVELGEKQFGEGDPRK 248

Qy      228  ASPELKSASAAVAHCKSKGKITK-----LALQYSLANKKEISSVLVGMSSVSQVEE 279
Db      249  RNPDFQGERPREILTAVDVLRLPIAKRYQATPAQLVLAWMKQNPVSVVIPGAKRPEQVSD 308

Qy      280  NVAATELSIGMDQETLSE 299
Db      309  NVQAL-DLHLSNEDYQIDE 327

RESULT 7
US-09-328-352-6079
; Sequence 6079, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6079
; LENGTH: 334
```

```
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6079

Query Match      18.3%; Score 296; DB 4; Length 334;
Best Local Similarity 27.0%; Pred. No. 1.8e-22;
Matches 93; Conservative 62; Mismatches 128; Indels 62; Gaps 10;

QY 3 KIELRAGNLTGLKVSAGVFGASPLGSGVFGPVAEDDAVATVREAFRLGINFFDTSPPYGGT 62
Db 1 KMKYRTLGQTGKVSALGCGMGSMFAYGASDDTQSIATLEKALDLGINFWDADMYGNG 60

QY 63 LSEKMLGKLKALQVPRSDYIVATKCG-RYKEG-----PDFSAERVRKSIDESL 110
Db 61 ANEVLLS---KYLEKHXDKVFLATKFGFRYKEDNLNPKNSLESYIDGSPWIKAVENS 117

QY 111 ERLQDYYVDILHCHDIEFGSLDOIYSETIPALOKLKQBGKTRFIGITGLPLDIFTYVLD 170
Db 118 RLNTDVIDLYAHRID---PNVPVEDTIGAMADLVKQGVRYLGLSEASAEITIRKAHAI 174

QY 171 VPPGTVDVILSYCHYGVNDSTLLDL-----LPYLKSKGVGVISASPLAMGLLTS- 219
Db 175 HPAAVQ-----HEYSLLTREFEQTHLTQIRELGISLVPPYSLRGLITNTLDVN 224

QY 220 -----QGPPEWHPASPELKSASKAHVAHCKSKGKITKLALQVSLANKEIS 265
Db 225 NLDENDFRQLPRYQG-DNW-----KXNOSLAQAFSEFAQSKNATAAALAWILAQGGDI 279

QY 266 SVLVGMSVSVQVEENVAATVLESIGMQETLSEVEAILPEPVKNL 310
Db 280 IPITRKIERLVENAGAV-----DLHLTAADLAELDALIARYPNM 320

RESULT 8
US-09-134-001C-3126
; Sequence 3126 Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3126
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3126

Query Match      18.0%; Score 291.5; DB 3; Length 314;
Best Local Similarity 27.1%; Pred. No. 4.9e-22;
Matches 86; Conservative 72; Mismatches 116; Indels 43; Gaps 12;

QY 12 TGLKVSAGVFGASPLGSGVFGPVAEDDAVATVREAFRLGINFFDTSPPYGGTSLKMLGKG 71
Db 20 SGIEISELGLGMSLGTDF-----KKAQPIESADTNGITFDITADIYDQGVNEISVGR 74

QY 72 LKALQVPRSDYIVATKCG--RYKEG---PDFSAERVRKSIDESLERLQDYYVDILHCHDI 126
Db 75 LKKYQ-NRDDIIVIGTKVGNRLTDDGHMTWDPKSKHKESVKSLKRLGLNHLDLQLHG- 132

QY 127 EFGSLDOIYSETIPALOKLKQBGKTRFIGITGLPLDIFTYVLDVPPGTVDVILSYCHY 186
Db 133 --GTIDDDPLDEITISAFDELKQBGYRAYGISSIRENVIDYLNK---SQIETLMS--QFN 185

QY 187 VNDSTLLDLPLYLKSKGVGVISASPLAMGLLT-----EQGPPEWHPASPELKS 235
```

```
Db 186 LIDNRPESLINDVHDQVKILARGPVFKGLTSSVDVIDEKFNGVLDY--TQDELGS- 242
QY 236 SKAAVAHCKSGKKITKLALQVSLANKEISSVLVGMSSVQVRENVAAVTELESIGMDOE 295
Db 243 ---TIASIKELSNLTALSPKYLTSHDAMGSIIVGASSVEQLEENVR--NYYKEISLQD- 296
QY 296 TLSEVEAILPEPVKNLTW 312
Db 297 ----IKSARNRVKDIIEY 309

RESULT 9
US-09-673-198-4
; Sequence 4 Application US/09673198
; Patent No. 6806076
; GENERAL INFORMATION:
; APPLICANT: MIYAKE, Koichiro; HASHIMOTO, Shinichi; MOTOVAMA Hiroaki;
; APPLICANT: OZAKI, Akio; SETO, Haruo; KUZAVAMA, Tomohisa; TAKAHASHI, Shunji
; TITLE OF INVENTION: A process for producing isoprenoid compounds by
; TITLE OF INVENTION: microorganisms and a method for screening compounds with
; TITLE OF INVENTION: antibiotic or weeding activity
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/673,198
; CURRENT FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: JP98/103101
; PRIOR FILING DATE: 1998-04-14
; PRIOR APPLICATION NUMBER: JP98/221910
; PRIOR FILING DATE: 1998-08-05
; PRIOR APPLICATION NUMBER: JP99/035739
; PRIOR FILING DATE: 1999-02-15
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-673-198-4

Query Match      18.0%; Score 291; DB 4; Length 348;
Best Local Similarity 28.4%; Pred. No. 6.5e-22;
Matches 94; Conservative 60; Mismatches 135; Indels 42; Gaps 11;

QY 4 IELRALGNTGLKVS-----AVGFGASPLGSGVFGPVAEDDAVATVREAFRLGINFFDTS 58
Db 25 MQYNPLGKTDLRVSRKLCGCMTFGEPRGNHAWTLPEESSRPIIKRALEGGINFDTANS 84

QY 59 YGGTSLSEKMLGKLKALQVPRSDYIVATK----CGRYKEGDFSAERVRKSIDESLERLQ 114
Db 85 YSDGSSEIIVGRLRDF-AREDDVVATKVFRVGDLPFG--LSRAQILRSIDDSLRRLG 141

QY 115 LDYVDILHCHDIEFGSLDOIYSETIPALOKLKQBGKTRFIGITGLPLDIFTYVLD-RVPP 173
Db 142 MDYVDILQIHRWDYNT---PIEETLEALNDVVYKAGARYIGASSMHASQFAQALELOKH 198

QY 174 GTVDVILSYCHYGVNDSTLLDLPLYLKSKGVGVISASPLAMGLLTTEQGPPEWHPASPEL 232
Db 199 GWAQFVSMQDHYNLIIYREEREMLPYCQEGVAVIPWSPLAGRLTR----PWGTTARL 254

QY 233 KS-----ASKAAVAH-----CKSKGKITKLALQVSLANKEISSVLVGMSSVS 275
Db 255 VSDEGVKNLYKESDENDADQIAERUTGVSEELGATRAQVALAWLLSKPGIAPIIGTSREE 314

QY 276 QVEENVAATVLESIGMDOETLSEVEAILPE 306
Db 315 QLDPELLNAV---DITLKPEQIAELETPYKP 341

RESULT 10
US-09-489-039A-12541
; Sequence 12541 Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
```

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12541
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12541

Query Match 17.1%; Score 277.5; DB 4; Length 364;
Best Local Similarity 28.6%; Pred. No. 1.8e-20;
Matches 98; Conservative 71; Mismatches 105; Indels 69; Gaps 15;

QY 4 TELRALGNTGLKVSAGVGASPLGSGVGFVGAEDDAVATRE-----APRLGINFPTSPYY 59
DB 47 MEYRCGRSGQLPAISLG---LWHPG---DETRVETSRQMLLHAFDLGITHFDLANNY 100
QY 60 G---GTLSEKMLGKGLKALQVP-RSDYIVATCGRYKEGDFPSAERVRK-----SIDESLE 111
DB 101 GPPPGS-AESNFGRIKLSLLPYRDELLIISTAGTMDGPGYDGMGSRKYLVASLNQSLK 159
QY 112 RLQLDYVDILHCHDIEFGSLDQIVSETIPALQKQEGKTRFIGHTGLPL-----161
DB 160 RMGLYVDIFVYHRPD---PQPLMETWALDHLVRQKALVVGISNYPPLAQAREAVKIL 216
QY 162 -DIFTYVLDVRVPGTVVILSYCHYGVNDSTLLDLLPYLKSQGVVISASPLAMGLLTFEQ 220
DB 217 NDLGTPCTIHPQR-----YSMPFERGVBEQ---LLDFLQTEGIGSIAPSPLAGGQLTDR 266
QY 221 -----GPPWHPASPE-LKSASKAAVAHCKSKGKITKALQOYSLAN 261
DB 267 YLNGIPADSRRAASSRFRLOPEQLTARLEKIKQLNRQA-----EARGQKLSQMALAWLURE 322
QY 262 KEISVLVGMSSVQSVQENVAATVLESGLMDQETLSEVEAIL 304
DB 323 EKVTSLVIGASKTAQLDD---AVGMLQNRHFTTECAIDAAIL 362

RESULT 11
US-09-724-623-86
; Sequence 86, Application US/09724623
; Patent No. 6476209
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Lubbers, Mark W
; APPLICANT: Dekker, James
; TITLE OF INVENTION: Polynucleotides, materials incorporating
; TITLE OF INVENTION: them, and methods for using them.
; FILE REFERENCE: 1048UI
; CURRENT APPLICATION NUMBER: US/09/724,623
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 86
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Lactobacillus rhamnosus
US-09-724-623-86

Query Match 17.1%; Score 277; DB 4; Length 334;
Best Local Similarity 27.7%; Pred. No. 1.8e-20;
Matches 96; Conservative 74; Mismatches 107; Indels 70; Gaps 18;

QY 3 KIELRALGNTGLKVSAGVGGA-SPLGVS--FGPVAEDDAVATREAPRLGINFPTSPYY 59
DB 11 KMPVRHAGKTGLMLPVISLGLWQHVGNDLPFGP-----RRSVILDAFGRGVHFDVANHY 65
QY 60 GG-----TLSEKMLGKGLKA-LQVPRSDYIVATKGRYK-----EGDFPSAERVRKSI 106

DB 66 GNGDREPFGSGSERLLGQILATDLKPYRDELIVSTKVG-YEIHFGPYGVGTSRKAIVIOGL 124
QY 107 DESLERLQLDYVDILHCHDIEFGSLDQI-VSETIPALQKQEGKTRFIGHTGLPL-----161
DB 125 NDSLKRLQLDYVDIYAHFDP---DVALSETVNALDQTVRDGKALYIGISNYDTKQTK 180
QY 162 -----DIFT-YVLDVRVPGTVVILSYCHYGVNDSTLLDLLPYLKSQGVVISASPLA 213
DB 181 EAIAMFKDLHTFPFLNQ-----YSYNNFNRTAET-SGLIDALKADGAGLIAYGPLS 230
QY 214 MGLLITE---QGPEW---HPASPELKSASKAAV-----AHCKSKGKITKALQY 257
DB 231 EGLLSDRYLKGPDPFTFKIHTPKATFAKGEAVVQKLNALNEIAH--DRDQTLQMALAW 288
QY 258 SLANKKEISSVLVGMSSVQSVQENVAATVLESGLMDQETLSEVEAIL 304
DB 289 LLRDPVVTSVIIGTTSVEHLQNLKAT---EHLATFAEEIQOIDDIL 332

RESULT 12
US-09-489-039A-9552
; Sequence 9552, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709-2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 9552
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9552

Query Match 16.9%; Score 274; DB 4; Length 350;
Best Local Similarity 27.7%; Pred. No. 4.1e-20;
Matches 101; Conservative 60; Mismatches 125; Indels 78; Gaps 17;

QY 9 LGNTGLKVSAGVGASPLGS---VFGPVAEDDAVA--TVREAFRLGINFPTSPYYGTL 63
DB 11 LGRSGLKVSRLCLGMSYGEPELPPQWSLDEKASRLIRQALEAGINFPTANIYSGS 70
QY 64 SEKMLGKGLKALQVPRSDYIVATKC-----GRYKEGDFPSAERVRKSIDESLERLQLDYV 118
DB 71 SEBITGKALREM-ARRDEIVVATKTFPPWRNSPNTGF-LSRKAIFQSIDDSLMRLGMDYV 128
QY 119 DILHCHDIEFGSLDQIVSETIPALQKQEGKTRFIGHTGLPLDIFT---YVLDVRVPGT 175
DB 129 DLFOIHRPDYST---PVEETMEALHDVVVKGAGKARYIGASSMEAWRFALKQHTAER--NGW 183
QY 176 VDVILSYCHYGVNDSTLL-----DLLPYLKSQGVVISASPLAMGLLTEQPPSWHPAS 229
DB 184 TRFITMQPOY-----NLLYREEREEMLPCLDEQGVGVIPWSPMARGRLTR-----DWSVTS 234
QY 230 -----PELKSASKAAVAHCKSKGKITKALQOYSLANKKEISSVL 268
DB 235 RRTQNDAPALKWYENAAALLDKPVIDVVVASIAEKUAVPR---AHVAIAWLLSKSVITAPI 290
QY 269 VGMSSVQSVQENVA-----VTELESGLMDQETLSEVEAILPEV-----KNLTWPSG 315
DB 291 IGATKPEHLSTAISALDPSLSDASITELEARYLPH-----PVDGIIPPLDPTPSPSPSA 346
QY 316 IQHN 319
DB 347 I-QN 349

RESULT 13
US-09-489-039A-12803
; Sequence 12803, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12803
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12803

Query Match 16.7%; Score 271; DB 4; Length 328;
Best Local Similarity 27.9%; Pred. No. 7.6e-20;
Matches 91; Conservative 60; Mismatches 133; Indels 42; Gaps 11;

QY 9 LGNTGLKVS-----AVGFGASPGVSGFVPAEDDAVATVREAFRLGNFFDTSPYGGTL 63
DB 10 LGDTALRVSRCLCGMTFGFGRHAWTLPESSRPLIQHATEGGINFFDTANSYSDGS 69
QY 64 SEKWLGLKALQVPRSDYVATK----CGRYKEGDFSAERVKKSIDESLERLQLDYVD 119
DB 70 SEEVIGSALRDF-ARRDEVVATKVHVQVGDLAG--LSRAQLRLSDDSLRLGMDYVD 126
QY 120 ILCHDIEFGSLDQIVSETIPALQKLQKQKTRFIGITGLPLDIFTVLD-RVPPGTVDV 178
DB 127 LLQIHRWDYTT----PIETIEALDEVVKAKARYIGASSMHAQFAQALALQOQNGWARF 183
QY 179 ILSYCHYG-VNDSTLLDLLPYLKSQGVISASPLAMGLLITEQGPPWHHPASPELKSASK 237
DB 184 VTQMDHYNLIVREENEMFLQCRNGVAVIPWSPFLARGRLTR---PWGETTARLVSEDF 239
QY 238 AAVAHCKSK-----GK-----KITKLALQVSLANKEISSVLVGMSSVSQVEEN 280
DB 240 GKSLYSTSEENDAQIAGKLADVAEELDASRAQVALAWLLSKPGVAAPFIIGPSRQEQDLDDL 299
QY 281 VAAVTELESGLMDQETLSEVEAILEP 306
DB 300 LQAV----DLTSPQEQIDKLEAPYQP 321

RESULT 14
US-09-252-991A-31535
; Sequence 31535, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31535
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31535

Query Match 16.3%; Score 264; DB 4; Length 358;
Best Local Similarity 28.3%; Pred. No. 4.8e-19;

Db 317 ATT-----ITLSDDTMSCLDEI-----WP 335

Search completed: November 13, 2005, 08:39:24
Job time : 29 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2005, 08:35:38 ; Search time 78 Seconds
(without alignments)
1711.187 Million cell updates/sec

Title: US-10-606-300-11

Perfect score: 1619

Sequence: 1 MTKIELRALGNTGLKVSANG.....VRAILEPVKNLTWPSTGHQ 319

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867879 seqs, 418409474 residues

Total number of hits satisfying chosen parameters: 1867879

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09D_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US10F_PUBCOMB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1619	100.0	319	15	US-10-240-136A-5
2	1619	100.0	319	17	US-10-606-300-11
3	1279	79.0	335	15	US-10-424-599-262305
4	1213	74.9	316	16	US-10-437-963-171001
5	1208.5	74.6	317	16	US-10-425-115-279898
6	1199.5	74.1	335	16	US-10-767-701-44783
7	1191.5	73.6	327	16	US-10-425-115-279897
8	908	56.1	229	15	US-10-425-114-69240
9	593.5	36.7	345	20	US-11-097-143-28023
10	544.5	33.6	342	20	US-11-097-143-40479
11	420.5	26.0	294	20	US-11-097-143-7146

12	385	23.8	328	14	US-10-156-761-8943	Sequence 8943, Ap
13	368.5	22.8	328	14	US-10-156-761-14813	Sequence 14813, A
14	337.5	20.8	306	15	US-10-369-493-23187	Sequence 23187, A
15	335	20.7	945	16	US-10-437-963-171001	Sequence 171001, A
16	329.5	20.4	304	15	US-10-282-122A-45565	Sequence 45565, A
17	329	20.3	341	15	US-10-425-114-46948	Sequence 46948, A
18	329	20.3	341	16	US-10-425-115-357386	Sequence 357386, A
19	329	20.3	341	16	US-10-739-930-8425	Sequence 8425, Ap
20	327	20.2	325	15	US-10-369-493-4875	Sequence 4875, Ap
21	322	19.9	308	15	US-10-369-493-7635	Sequence 7635, Ap
22	322	19.9	329	15	US-10-282-122A-57831	Sequence 57831, A
23	316	19.5	307	15	US-10-369-493-16460	Sequence 16460, A
24	316	19.5	346	15	US-10-282-122A-55730	Sequence 55730, A
25	313	19.3	326	9	US-09-815-242-10186	Sequence 10186, A
26	313	19.3	326	15	US-10-369-493-862	Sequence 862, App
27	313	19.3	326	15	US-10-282-122A-43244	Sequence 43244, A
28	310.5	19.2	302	15	US-10-282-122A-71604	Sequence 71604, A
29	310.5	19.2	331	14	US-10-156-761-8701	Sequence 8701, Ap
30	310	19.1	311	15	US-10-282-122A-45545	Sequence 45545, A
31	309.5	19.1	397	15	US-10-424-599-156741	Sequence 156741, A
32	308	19.0	329	15	US-10-424-599-156752	Sequence 156752, A
33	307.5	19.0	311	15	US-10-369-493-19379	Sequence 19379, A
34	306.5	18.9	332	15	US-10-282-122A-74948	Sequence 74948, A
35	305.5	18.9	323	9	US-09-815-242-11846	Sequence 11846, A
36	305.5	18.9	332	15	US-10-282-122A-75573	Sequence 75573, A
37	302	18.7	305	15	US-10-282-122A-45821	Sequence 45821, A
38	301.5	18.6	339	15	US-10-369-493-20165	Sequence 20165, A
39	300.5	18.6	320	15	US-10-369-493-19397	Sequence 19397, A
40	297.5	18.4	346	15	US-10-424-599-228878	Sequence 228878, A
41	297.5	18.4	357	15	US-10-425-114-55914	Sequence 55914, A
42	296.5	18.3	315	15	US-10-282-122A-57996	Sequence 57996, A
43	295	18.2	344	15	US-10-369-493-12039	Sequence 12039, A
44	294.5	18.2	314	15	US-10-369-493-20010	Sequence 20010, A
45	294.5	18.2	341	16	US-10-739-930-10839	Sequence 10839, A

ALIGNMENTS

RESULT 1
US-10-240-136A-5
; Sequence 5, Application US/10240136A
; Publication No. US20040053235A1
; GENERAL INFORMATION:
; APPLICANT: Ascorbex Limited
; TITLE OF INVENTION: Gene Sequence
; FILE REFERENCE: JPD/P100606US-PCT
; CURRENT APPLICATION NUMBER: US/10/240,136A
; CURRENT FILING DATE: 2003-07-14
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; US-10-240-136A-5

Query Match	100.0%	Score	1619	DB	15	Length	319
Best Local Similarity	100.0%	Pred. No.	1.3e-149				
Matches	319	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	1	MTKIELRALGNTGLKVSANGPVGASPLGSPGVAEDDAVATVRAFRALGINFFDTSPPYG	60				
Db	1	MTKIELRALGNTGLKVSANGPVGASPLGSPGVAEDDAVATVRAFRALGINFFDTSPPYG	60				
Qy	61	GTLSEKMLGKGLKALQVPRSDYIVATKGRYKEGDFSAERVKSIDSLERLQLDYVDI	120				
Db	61	GTLSEKMLGKGLKALQVPRSDYIVATKGRYKEGDFSAERVKSIDSLERLQLDYVDI	120				
Qy	121	LHCHDIFSGSLDQIVSETIPALQKKGKTRPTGITGLPLDITFYVLDVRPPGTVVIL	180				
Db	121	LHCHDIFSGSLDQIVSETIPALQKKGKTRPTGITGLPLDITFYVLDVRPPGTVVIL	180				

Db 121 LHCHDIEFGSLDQIVSETIPALQKLKQEGKTRFIGITGLPLDIFTYVLDVRVPPGTVDVIL 180
Qy 181 SYCHGVNDSTLLDLLPYLKSQGVVISASPLAMGLLTEQGGPPPEWHHPASPELKSAAV 240
Db 181 SYCHGVNDSTLLDLLPYLKSQGVVISASPLAMGLLTEQGGPPPEWHHPASPELKSAAV 240
Qy 241 AHCKSKGKIKTKALQYSLANKEISSVLVGMSSVSQVEENVAATELESGLMDQETLSEV 300
Db 241 AHCKSKGKIKTKALQYSLANKEISSVLVGMSSVSQVEENVAATELESGLMDQETLSEV 300
Qy 301 EAILPEPVKNLTWPSGIHQN 319
Db 301 EAILPEPVKNLTWPSGIHQN 319

RESULT 2

US-10-606-300-11
; Sequence 11, Application US/10606300
; Publication No. US20050019879A1
; GENERAL INFORMATION:
; APPLICANT: Porto, Danilo
; APPLICANT: Sauer, Michael
; TITLE OF INVENTION: Ascorbic Acid Production from Yeast
; FILE REFERENCE: 2028.594000
; CURRENT APPLICATION NUMBER: US/10/606,300
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-606-300-11

Query Match 100.0%; Score 1619; DB 17; Length 319;
Best Local Similarity 100.0%; Pred. No. 1.3e-149; Indels 0; Gaps 0;
Matches 319; Conservative 0; Mismatches 0;

Qy 1 MTKIELRALGNTGLKVSAGFGASPLGSVFGPVAEDDAVATVREAFRLGINFFDTSPPYG 60
Db 1 MTKIELRALGNTGLKVSAGFGASPLGSVFGPVAEDDAVATVREAFRLGINFFDTSPPYG 60
Qy 61 GTLSEKMLGKGLKALQVPRSDYIVATKCGRYKEGDFSAERVKSIDESLERLQDYVDI 120
Db 61 GTLSEKMLGKGLKALQVPRSDYIVATKCGRYKEGDFSAERVKSIDESLERLQDYVDI 120
Qy 121 LHCHDIEFGSLDQIVSETIPALQKLKQEGKTRFIGITGLPLDIFTYVLDVRVPPGTVDVIL 180
Db 121 LHCHDIEFGSLDQIVSETIPALQKLKQEGKTRFIGITGLPLDIFTYVLDVRVPPGTVDVIL 180
Qy 181 SYCHGVNDSTLLDLLPYLKSQGVVISASPLAMGLLTEQGGPPPEWHHPASPELKSAAV 240
Db 181 SYCHGVNDSTLLDLLPYLKSQGVVISASPLAMGLLTEQGGPPPEWHHPASPELKSAAV 240
Qy 241 AHCKSKGKIKTKALQYSLANKEISSVLVGMSSVSQVEENVAATELESGLMDQETLSEV 300
Db 241 AHCKSKGKIKTKALQYSLANKEISSVLVGMSSVSQVEENVAATELESGLMDQETLSEV 300
Qy 301 EAILPEPVKNLTWPSGIHQN 319
Db 301 EAILPEPVKNLTWPSGIHQN 319

RESULT 3

US-10-424-599-262305
; Sequence 262305, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 262305
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_78884C.1.pep
US-10-424-599-262305

Query Match 79.0%; Score 1279; DB 15; Length 335;
Best Local Similarity 76.8%; Pred. No. 2.8e-116;
Matches 245; Conservative 37; Mismatches 37; Indels 0; Gaps 0;

Qy 1 MTKIELRALGNTGLKVSAGFGASPLGSVFGPVAEDDAVATVREAFRLGINFFDTSPPYG 60
Db 17 LREMELRELGRITGUKLSTVGFASPLGNVFGDVEEQANASVRLAFQSGINFFDTSPPYG 76
Qy 61 GTLSEKMLGKGLKALQVPRSDYIVATKCGRYKEGDFSAERVKSIDESLERLQDYVDI 120
Db 77 GTLSEKVLGKALKALGAPRNSYVATKCGRYKEGDFSAERVTRISIESLERLQDYVDI 136
Qy 121 LHCHDIEFGSLDQIVSETIPALQKLKQEGKTRFIGITGLPLDIFTYVLDVRVPPGTVDVIL 180
Db 137 LQCHDIEFGSLDQVNETIPALVKLKEAGKARFIGITGLPLGIFSYVLDVRVPPGTVDVIL 196
Qy 181 SYCHGVNDSTLLDLLPYLKSQGVVISASPLAMGLLTEQGGPPPEWHHPASPELKSAAV 240
Db 197 SYCHGVNDSTLSGELVPYLKTKGVIINASPLSGLLTESGPPPEWHHPASPELKSACQAAA 256
Qy 241 AHCKSKGKIKTKALQYSLANKEISSVLVGMSSVSQVEENVAATELESGLMDQETLSEV 300
Db 257 THCKEKGKNIKALQYSLANKEITSVLVGMKSVEQVEENVAARELATSGIDBETLSEV 316
Qy 301 EAILPEPVKNLTWPSGIHQN 319
Db 317 EAILPKVKNQSWPSGIQOS 335

RESULT 4

US-10-437-963-113940
; Sequence 113940, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 113940
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_17680C.1.pep
US-10-437-963-113940

Query Match 74.9%; Score 1213; DB 16; Length 316;
Best Local Similarity 74.3%; Pred. No. 7.6e-110;
Matches 234; Conservative 29; Mismatches 52; Indels 0; Gaps 0;

[illegible]

RESULT 5

```

US-10-425-115-279898
; Sequence 279898, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 279898
; LENGTH: 317
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MFT4577 186846C.1.pep

```

Query Match	74.6%	Score	1208.5	DB	16	Length	317
Best Local Similarity	74.7%	Pred. NO.	2.1e-109				
Matches	236	Conservative	29	Mismatches	50	Indels	1
						Gaps	1

Qy	4	I E L R A L G N T G L K V S A V G F G A S P L G S V F G P V A E D D A V A T V R E A F R L G I N F F D T S P Y Y G G T L	63
Db	1	M E L R E L G G T G L R V S A V G F G A S P L G N V F G D P V R A A V R R A L D G I N F F D T S P Y Y G G T I	60
Qy	64	S E K W L K G K I K A L Q V R P S D Y I V A T K C G R Y K - E G F O F S A E R V R K S I D E S L E R L Q L D Y V D I L H	122
Db	61	S E S V L G D C L R H A A V P R D R V V A T K C G R Y K D E G F D F S A N R V T R S I D E S L A R L G L D Y V D I L H	120
Qy	123	C H D I E P G S L D Q I V S E T T I P A L O K L K O E G K T R I G I T G L P L D I F T V Y L D R V P P G T V D V I L S Y	182
Db	121	A H D I E F T H L E Q I V N E T I P A L O K I K E N G K A R T I G I T G L P L S I Y P Y V L D R V A P G S V D I L S Y	180
Qy	183	C H Y G V N D S T L D L L P Y L K S K G V G I S A S P L A M G L L T E Q G P P E W H P A S P E L K S A K A A V A H	242
Db	181	C H Y G I N D T S L V D L L P Y L K S K G V I S A S P L S M G L L T D N G P P E W H P A P E E L K S A C K A A A D H	240
Qy	243	C K S K G K I T K L A I O V S L A N K E I S S V L V G M S S V S Q V E E N V A A V T E L S L G M D Q E T L S E V A	302
Db	241	C R K K G S I T K L A M Q V S L M N N E I S T V L V G M N S L E Q V E E N V A A A L S T S G I D E L M R E V E A	300
Qy	303	I L E P V K N L T W P S G I H Q	318

```

Db          301 ILEPVKLTWPSGIQQ 316
|||||
RESULT 6
US-10-767-701-44783
; Sequence 44783, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 44783
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C59142_1.pep
US-10-767-701-44783

```

Query Match	74.1%	Score 1199.5;	DB 16;	Length 335;
Best Local Similarity	73.7%;	Pred. No. 1.7e-108;		
Matches 233;	Conservative 32;	Mismatches 50;	Indels 1;	Gaps 1;

Qy	4	IELRALGNTGLKYSVAGFGASPLGSGVFGPVAEDDAVATVREAFLRGINFDDTSPYVGGTL	63
Db	19	MEURELGGTGLRVSAGVFGASPLGNVFGDPDRDARAARRALDGLGNFDDTSPYGGTV	78
Qy	64	SEKMLGKGLKALQVPRSDYIVATKCCGRYK-EGPDFSAERVKGISDESLERLQLDYYVDILH	122
Db	79	SESVLGDCLRHAAPRDRVVVATKCGRYKDEGPDFSADRVTRSIDESLARLGLDYYVDILH	138
Qy	123	CHIEFSGSLDQIVSETIPALOKLKQEGKTRFIGITGLPLDIFTYVLDRVPGTVDWILSY	182
Db	139	AHDIEFTHLDQIVNETIPALQKIKESGARFIGITGLPLSIYPVLDRAVPGSDVLILSY	198
Qy	183	CHYGVDNSTLLDLLPYLKSXGVGVI SAS PLAMGLL TEQGPPEWHPASPELKSASKAAVAH	242
Db	199	CHYGINDTSVLDDLPLYLKSXGVGVI SASPLSMGLLTDNGPPEWHPAEQULKSACRAAADH	258
Qy	243	CKSGKKIKTKLAQYSLANKBEISVLVGMSSSVQVEENVAAVTELSGLGMDQETLSVEEA	302
Db	259	CRKGRSITKLAMQYSLMWNNEISTVLVGMSSSVQVEENVAALSTSGIDELLREVEA	318
Qy	303	ILPFPVKLWTWPSGIHQ	318
Db	319	ILPFPVKLWTWPSGIOO	334

```

RESULT 7
US-10-425-115-279897
; Sequence 279897, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 279897
; LENGTH: 327
; TYPE: PRT

```

```
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_186845C.1.pap
US-10-425-115-279897

Query Match      73.6%; Score 1191.5; DB 16; Length 327;
Best Local Similarity 72.4%; Pred. No. 1e-107;
Matches 236; Conservative 28; Mismatches 51; Indels 11; Gaps 2;

QY 4 IELRALGNTGLKVSAGVGFASPLGVSFGVPAEDDVAATVREAFRLGINFEDTSPYGGTL 63
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MELRELGGTGLRVSAGVGFASPLGVSFGVPRDTRAAVRRALDGINFEDTSPYGGTI 60

QY 64 SEKMLGKGLKALQVPRSDYIVATKCGRYK-EGFDFSARVRKSIDESLERLQLDYVILH 122
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 SESVLGDCILRHAAVPRDTRVATKCGRYKDEGDFSANRVTRSIDESLARLGLDYVILH 120

QY 123 CHDIEFGSDQIVSETIPALQKQEGKTRFIGITGLPLDIFTYVLDVPPCTVDVILSY 182
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 AHDIETFLHLEQIVNETIPALQKIKENGKARFIGITGLPLSIYPYVLDRAVPGSDVILSY 180

QY 183 CHYGVNDSTLLDLLPYLKSQGVVISASPLAMGLLLEQGPPEWHHPASPEL----- 232
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 CHYGVNDSTLLDLLPYLKSQGVVISASPLAMGLLLEQGPPEWHHPASPELKMPEFYLENE 240

QY 233 KSASKAQVAHCKSGKKTITKALQYSLANKEISSVLVGMSSVSVQVEENVAATELESIGM 292
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 KSASKAQVAHCKSGKKTITKALQYSLANKEISTVLVGMNSLEQVEENVAATELESIGI 300

QY 293 DOETLSEVEAILEPVKNLTWPSGIHQ 318
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 DDELMREVEAILEPVKNLTWPSGIHQ 326

RESULT 8
US-10-425-114-69240
; Sequence 69240, Application US/10425114
; Publication NO. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; PRIOR FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 69240
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73186H07_FLI.pap
US-10-425-114-69240

Query Match      56.1%; Score 908; DB 15; Length 229;
Best Local Similarity 77.0%; Pred. No. 3.6e-80;
Matches 174; Conservative 22; Mismatches 30; Indels 0; Gaps 0;

QY 93 EGFDFSARVRKSIDESLERLQLDYVDIILHCHDIEFGSDQIVSETIPALQKQEGKTR 152
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3 EGFDFSANRVTRSIDESLARLGLDYVDIILHCHDIEFTHLEQIVNETIPALQKIKENGKAR 62

QY 153 FIGITGLPLDIFTYVLDVPPCTVDVILSYCHYGVNDSTLLDLLPYLKSQGVVISASPL 212
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 63 FIGITGLPLSIYPYVLDRAVPGSDVILSYCHYGVNDSTLLDLLPYLKSQGVVISASPL 122

QY 213 ANGLLLEQGPPEWHHPASPELKSASKAQVAHCKSGKKTITKALQYSLANKEISSVLVGM 272
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```



```

; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8943
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8943

Query Match          23.8%; Score 385; DB 14; Length 328;
Best Local Similarity 29.5%; Pred. No. 1.1e-28;
Matches 101; Conservative 68; Mismatches 127; Indels 46; Gaps 8;

Qy 4 IEIRALGNLTGLKVSAGVFGASPLGSGVGPVAEDDAVATVREAPRLGINFFDTSPYGGTLL 63
Db 1 VKARPLGRTSVRVSTLGFGAAPLGNLYGPIDDCQAATLKAADVGVVRYDTAPHYGLGL 60
Qy 64 SERMLHGKLKALQVPRSDYIVATKGRYKE-----GF-----DPSAE 100
Db 61 SERRLGDAL--AHRPRSEFTISTKVGRLLEPHAPAGTSGDLTAGGFVDPDTLVRRPDYSRD 118
Qy 101 RVKSIDESLERLDYVDILHCHDIEFGSLDQIVSETIPALQKLKQEGKTRFTGIICLP 160
Db 119 GVLRSEGLSNRLRLDHDIVVYHDDP--DHLDAALDHALPAUTALRDQG---VITGAVGV 174
Qy 161 LDIFTYVLDRVPPGTVDVILSYCHGVNDSTLLDLLPYLKSKGVGVTASPLAMGLLLEQ 220
Db 175 MNVAVPLRVVAEAEADVAVAGRWTLDRTARELLDCAERGVAVVAAPFNGLLSRP 234
Qy 221 GPP-----EWHPASPELKSASKAAVHCKSKGKKITKLALQYSLANKEISSVLVGMSSVS 275
Db 235 HPANDATFDYGPASESALRRARLLAGVSGRHGTALPHAALRFLRPDRPSVACVAVGFRFSE 294
Qy 276 QVENVA-AVTEESLGMDQETLSEVRAILEPVKNLTPWPGI 316
Db 295 EVVSAARWATDITG-----EAWLDLDDTAATWAARI 325

```

```

RESULT 13
US-10-156-761-14813
; Sequence 14813, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14813
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14813

```

Query Match	22.8%	Score 368.5;	DB 14;	Length 328;
Best Local Similarity	34.1%	Pred. NO. 4.5e-27;		
Matches 101; Conservative	49;	Mismatches 115;	Indels 31;	Gaps 7;

Qy	9	LGNTGLKVSAGFGASPLGVSFGPVAEDDAVATVREAFRLGINFFDTSPPYGGTSLSEKML	68
Db	4	LGRSGEVSELSFGAAGTGNLFTAVTDEQASQAVHAATSGIRYFDTAPHYGLGLSERRL	63
Qy	69	GKGLKALQVPRSDYIVATKCGRYKE-----GF-----DPSAERVKSKID	107
Db	64	GAALR---EHPTATVTVTKVGRLEPTDAGDDLDAGFAVPATHHRVWDFSDGVRRALT	121
Qy	108	ESLERLQDYVDYIILHCHDIEFGSLDQIYSETIPALQKLQKQEGKTRFIGITGLPLDIFTVV	167
Db	122	ASLERLGLDRVDVVLVHDPD-DHAEQAPEGGCPALEKLRSEG--VVGAIAGMNQTAML	177
Qy	168	LDRVPPGTVDVILSYCHGVNDS-TLLDLLPYLKSQKGVGVIASPLAMGLLITEQGP--	223
Db	178	TRFYRETDVDVVLACAGRYTLLDORALADLLPAAADRGTSVWIGGAFNSGLLADPKPGATY	237
Qy	224	EWHPASPELKASAAVAHCKSGKKITKLALQYSLANKEISSVLVGMSSVSQVEE	279
Db	238	NYAAAPPELLDRALRLKXVAERHGTTLRGAALAFCAAHPAVASVLVGARSPYEVRD	293
RESULT 14			
US-10-369-493-23187			
; Sequence 23187, Application US/10369493			
; Publication No. US20030233675A1			
; GENERAL INFORMATION:			
; APPLICANT: Cao, Yongwei			
; APPLICANT: Hinkle, Gregory J.			
; APPLICANT: Slater, Steven C.			
; APPLICANT: Goldman, Barry S.			
; APPLICANT: Chen, Xianfeng			
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF			
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES			
; FILE REFERENCE: 38-10(52052)B			
; CURRENT APPLICATION NUMBER: US/10/369,493			
; CURRENT FILING DATE: 2003-02-28			
; PRIOR APPLICATION NUMBER: US 60/360,039			
; PRIOR FILING DATE: 2002-02-21			
; NUMBER OF SEQ ID NOS: 47374			
; SEQ ID NO 23187			
; LENGTH: 306			
; TYPE: PRT			
; ORGANISM: Bacillus subtilis			
US-10-369-493-23187			
Query Match 20.8%; Score 337.5; DB 15; Length 306;			
Best Local Similarity 31.3%; Pred. No. 4.4e-24;			
Matches 100; Conservative 64; Mismatches 108; Indels 47; Gaps 12;			
Qy	7	RALGNTGLKVSAGFGASPLGVSFGPVAEDDAVATVREAFRLGINFFDTSPPYGGTSLSEK	66
Db	4	RKLGTSDLDISVGLGCMSLGT-----EKNKALSILDEAIELGINFLYDTADLYDRGNEE	58
Qy	67	MLGKGLKALQVPRSDYIVATKCG-RYKEG-----PDFSAERVKSKIDESLERLQDYVDI	120
Db	59	IVG---DAIQNRHDIILATAGNRWDDSGEYWDPSKAYIKEAVKKSILRLKTDYIDL	115
Qy	121	LHCHDIEFGSLDQIYSETIPALQKLQKQEGKTRFIGITGL-PLDIFTYVLDVRPPGVVDVI	179
Db	116	YQLHG---GTIEDNIDETIEAPEELKQEGVIRYGISSIRPNVIKEYVKS-----NIV	166
Qy	180	LSYCHGVNDS-TLLDLLPYLKSQKGVGVIASPLAMGLLITEQGPPEHPAS-----	229
Db	167	SIMMQFSLFDRRPEWPLLEHQLISVVARGVAGKLLTEK-PLQOASESMKQNGYLSYS	225
Qy	230	-PELKASKA--AAVAHCKSGKKITKLALQYSLANKEISSVLVGMSSVSQVEENVA--	283
Db	226	FEELTNARKAMEEVA PDL-----MTEKSIQLYLLAQPAVASVITGASKIEQLRENIQAANA	281
Qy	284	--VTELESGLMDQETLSEV	300
Db	282	RLTTEEEIKALQSHTKQDI	300

Search completed: November 13, 2005, 08:40:50
Job time : 79 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2005, 08:30:22 ; Search time 16 Seconds
(without alignments)
1918.322 Million cell updates/sec

Title: US-10-606-300-11
Perfect score: 1619
Sequence: 1 MTKIELRALGNTGLKVSAGV.....VEAILEPVKNLTWPSGIHQN 319

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1619	100.0	319	2 T04984	hypothetical prote
2	503	31.1	439	2 T28841	hypothetical prote
3	366	22.6	329	2 JC2405	D-threo-aldoase 1-d
4	362.5	22.4	297	2 C83776	oxidoreductase Bhl
5	358	22.1	335	2 S52890	hypothetical prote
6	346.5	21.4	329	2 AF2990	oxidoreductase Atu
7	346.5	21.4	329	2 C98293	D-threo-aldoase 1-d
8	337.5	20.8	306	2 H69966	conserved hypothet
9	317	19.6	351	2 AE2979	aldo/keto reductas
10	317	19.6	351	2 H98303	hypothetical prote
11	314	19.4	326	2 H90938	hypothetical prote
12	314	19.4	326	2 D85787	hypothetical prote
13	313	19.3	326	2 C64937	hypothetical prote
14	305.5	18.9	323	2 H83427	probable oxidoredu
15	305.5	18.9	332	2 AB0808	probable ion-chann
16	305	18.8	319	2 H72391	K+ channel, beta s
17	304.5	18.8	338	2 G95357	probable [imported
18	301.5	18.6	332	2 D98252	general stress pro
19	301.5	18.6	332	2 AH3033	aldo-keto reductas
20	298.5	18.4	353	2 AH3156	oxidoreductase moc
21	298.5	18.4	353	2 A98131	moca protein limpo
22	292	18.0	310	2 D69646	myo-inositol catab
23	291	18.0	324	2 A90688	probable NAD(P)H-d
24	291	18.0	324	2 E85538	probable oxidoredu
25	291	18.0	348	2 C64771	probable ion chann
26	290	17.9	319	2 T35337	aldo/keto reductas
27	289	17.9	369	2 C97390	probable reductase
28	287.5	17.8	346	2 E91114	aldo-keto reductas
29	287	17.7	333	2 AD2608	aldo-keto reductas

30	286.5	17.7	327	2	E75296	aldo/keto reductas
31	286.5	17.7	338	2	AI3183	aldo/keto reductas
32	284	17.5	301	2	B86966	probable oxidoredu
33	284	17.5	324	2	AE0554	probable oxidoredu
34	283	17.5	310	2	B69755	ion channel homolo
35	281.5	17.4	346	2	E85959	probable reductase
36	277.5	17.1	346	2	G85086	hypothetical prote
37	276.5	17.1	329	2	B82644	sugar-phosphate de
38	276.5	17.1	362	2	D82644	probable oxidoredu
39	276	17.0	353	2	F95403	probable aldoketo
40	274.5	17.0	331	2	B95902	probable aldoketo
41	274.5	17.0	354	2	D90187	oxidoreductase (im
42	271.5	16.8	379	2	E72284	oxidoreductase, al
43	270.5	16.7	340	2	B96632	hypothetical prote
44	270	16.7	329	2	AI0341	probable aldo/keto
45	268.5	16.6	315	2	A75289	probable potassium

ALIGNMENTS

RESULT 1

T04984
hypothetical protein T16L1.160 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C:Accession: T04984
R:Bevan, M.; Obermaier, B.; Deutschenbaur, S.; Piravandi, E.; Hoheisel, J.; Mewes, H.W.;
submitted to the Protein Sequence Database, November 1998
A:Reference number: Z15393
A:Accession: T04984
A:Molecule type: DNA
A:Residues: 1-319 <BEV>
A:Cross-references: UNIPROT:O81884; EMBL:AL031394
A:Experimental source: cultivar Columbia; BAC clone T16L1
C:Genetics:
A:Map position: 4
A:Introns: 57/2; 133/3; 233/3; 276/3
A:Note: T16L1.160
C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match 100.0%; Score 1619; DB 2; Length 319;
Best Local Similarity 100.0%; Pred. No. 4.7e-114;
Matches 319; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MTKIELRALGNTGLKVSAGVGFASPLGSGVFGFVAEDDAVATVREAFRLGINFFDTSPPYVG	60
Db	1	MTKIELRALGNTGLKVSAGVGFASPLGSGVFGFVAEDDAVATVREAFRLGINFFDTSPPYVG	60
Qy	61	GTLSEKMLGKGLKALQVPRSDYIVATKGRYKEGDFSAERVRKSIDESLERLQLDYVDI	120
Db	61	GTLSEKMLGKGLKALQVPRSDYIVATKGRYKEGDFSAERVRKSIDESLERLQLDYVDI	120
Qy	121	LHCHDIEFGSLDQIVSETIPALQKLKQSGKTRFGITGLPLDIPFYVLDVRVPPGVTVIL	180
Db	121	LHCHDIEFGSLDQIVSETIPALQKLKQSGKTRFGITGLPLDIPFYVLDVRVPPGVTVIL	180
Qy	181	SYCHGVNDSTLLDLPYLKSGKGVISASPLAMGLLTEQGPPEWHSPASPELKSAAV	240
Db	181	SYCHGVNDSTLLDLPYLKSGKGVISASPLAMGLLTEQGPPEWHSPASPELKSAAV	240
Qy	241	AHCKSKGKKITKALQVSLANKEISSVLVGMSSVSQVEENVAATELSLGMQDTLSV	300
Db	241	AHCKSKGKKITKALQVSLANKEISSVLVGMSSVSQVEENVAATELSLGMQDTLSV	300
Qy	301	EAILEPVKNLTWPSGIHQN 319	
Db	301	EAILEPVKNLTWPSGIHQN 319	

RESULT 2
T28841
hypothetical protein F37C12.12 - Caenorhabditis elegans


```
64 SRKMLGKGLAKQVPRSDYIVATKCGRYKE-----GF-----DPSAERV 103
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db SERRVGDFLQ--EKRPDEFVLSTKVGRILKPABEAGVPDYGFVDALPFIVEYDSYDGIM 118

      .
104 KSIDSLEKLQLDYDIILCHIDIFBSLOO-----IVSET-IPALOKLKOEKTRRFI 154
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db RSHELSLARLGSLGVLDLYVHDLEATTLCSEAVRRHFGIFTESGIEALHELKAKE---I 175

      .
155 GITGLPLDITFTTVLRVPGTVDVIUSYCHYGNDSTLLD-----LLPYLSKGCVGIS 208
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db GAFGLGVNEVPACLNMLEIDEICILLAGRY----TLDRSAAARLLRCRAETGTSLVI 230

      .
209 ASPLAMGLLTTEOGPP-----EWIPASPELSKSKAAVAHCKSGKKKTKLALQVSLANKEI 264
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db GGVFNSGIILATCAKPGATFNYNAEVPEVMERVCAMEAHGAHGVALAAAHPFLQNTDV 290
```

```

QY      265  SSVLVCMSSVSQVEENVA-----AVTELSLGHMDQETLSE 299
       :|||::: |||: | | | | |
DB      291  ASVLIQTAKPDSLRRRLSIFETAVPGAAWAGFDTLALED 329

RESULT 7
C98293 D-2-threo-aldoase 1-dehydrogenase (EC 1.1.1.122) [imported] - Agrobacterium tumefaciens
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: C98293
A:Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qureshi, B.; Li, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Mar Science 294, 2323-2328, 2001
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; UID:21608551; PMID:11743194
A:Accession: C98293
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-329 <KUR>
A:Cross-references: UNIPROT:Q8UAU7; GB:AE007870; PIDN:AAK89869.1; PID:gl5159812
C:Genetics:
A:Gene: AGR L 2607

```

Query Match	21.4%;	Score 346.5;	DB 2;	Length 329;
Best Local Similarity	30.1%;	Pred. NO. 1.6e-18;		
Matches 102;	Conservative 64;	Mismatches 120;	Indels 53;	Gaps 10;
QY	4	IEURLGNTGLKVSAGVFGASPLGVSFGPVAEDDAVATVREARFLGNINFPDTSFYCGTL	63	
Db	1	MKTKLGRTALELSLSFGAAGTGNLYRSVRSDAMATLQTAWDAGIRVEDTAPYQOGL	60	
QY	64	SEKWLKGKLGALQVPRSDYIVATKCRYKE-----GF-----DPSAERVR	103	
Db	61	SEERVGDPLQ--EKPRDEFVLSTKVRILKPABEAVTPDYGFDALPFIVEDYSYDGM	118	
QY	104	KSIDESLERLQLDYVDILCHDIEFGSLDQ-----IVSET-IPALOKLKGEGKTRFI	154	
Db	119	RSHELSTARLGLSGVDILVHDDLEATLGEAYVRHHGIFTESGIEALHKKAGE--	175	
QY	155	GITGLPLDIFTYVLDVRPPGTVDVILSYCHYGVNDSTLLD-----LLPYLKSKGVGVIS	208	
Db	176	GAFGLGVNEVPACLNLMETDEIDCILLAGRY-----TLDRSAARLLGCAETGTSLVI	230	
QY	209	ASPLAMGLLTQGGP-----EWHPASPELSKASKAAVAHCKSKGKITTKLALOYLSANKEI	264	

```

265 SSVLCMSSVSQVEENVA----AVTELSLGMDOETLSE 299
      :|||: : : : :|
291 ASVLIGTAKPDSLRRLNLSIFETAVPGAAWAGFTLALD 329

```

[illegible]

C;Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query March 19.6%; Score 317; DB 2; Length 351;
Best Local Similarity 29.8%; Pred. NO. 2.9e-16;
Matches 101; Conservative 60; Mismatches 128; Indels 50; Gaps 12;

Qy 4 IEURLGNTGLKVSAGVGFASPLGSVGFPVAEDDAVATVREAPRLGINFDTSPYGGTL 63
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 2 VDVRYLGRSALKVSPLSLGTMMFG---GPTPDDVAVRIIDKAREQQINFITDADVHDGK 58

Qy 64 SEKNMKGKLKALQVPDSDIYIVTK-CGRYEKEGDF-----SAERVKSIDSLELRQLDYVD 119
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 59 SEBVGWGRIKA----SRDHWYLAIFKFNSTHTKGNLGHSRKKWIETVENSRLRNTDYID 115

```

Qy 120 ILCHDIEFGSLDQIVSETIPALQKLQKQKTRPIGTT---GUPLDIFTYVLDRVPPGTV 176
Db 115 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116 ILYFHRAVF---DAPLEEPVRAIADLRAGKLRVFGVSNFRGWRIBIASHLADQL--GID 170
Qy 177 DVILSYCHYG-VNDSTLIDLPLLYLXSGVGVISASPLAMGLLTQGGPP-----223
Db 171 RPVASQPLYNIIVNRTAEQPLPAANHLYGLGVSVSPLARGVLTGKYQFGQPGADTRVGR 230
Qy 224 -----EWHPASPELKSASKAVAHACKSGKRRKITKLALQYLSIANKEISSVLVGMSSVS 275
Db 231 GDKRVLETEWRPESVEI--AQKVA-AHAASKGVSAADPALAWVLNNKFVTAATIGPTEE 287
Qy 276 QVEENVAAVTLESLGMDQETLSEVAILEPVRKHLTWPS 314
Db 288 HMNSYVRA-----LDVRIDAEDALVDLSVTTGHPS 318

RESULT 11
H90938
hypothetical protein ECs2480 [imported] - Escherichia coli (strain O157:H7, substrain R157)
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C/Accession: H90938
R:Hayaashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome analysis
A/Reference number: A99629; MUID:21156231; PMID:11258796
A/Accession: H90938
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1,326 <HAY>
A/Cross-references: UNIPROT:Q8XDV3; GB:BA000007; PIDN:BA035903.1; PID:g13361947; GSPDB:G13361947
A/Experimental source: strain O157:H7, substrain R1MD 0509952

```

Query Match 19.4%; Score 314; DB 2; Length 326;
Best Local Similarity 30.0%; Pred. No. 4.4e-16;
Matches 103; Conservative 46; Mismatches 128; Indels 66; Gaps 12;

Qy		1	MTKIELRALGNTGLKVSAGVGFSGSPLGSVGPVAED-----AVATVREAFRLGINFFPT	55
Dd		1	MKKI---PLGTDTITLSRMGLGTWAGG--GPWANGDLRQCIDTILEAHRCGINLIDT	55
	:	:	:	:
Qy		56	SPIYGGTSLSEKNMLKGKLALQPVRSDYIVATKCQ-----RYKEGFDPS	98
Dd		56	APGVNFCNSEVIGVALKKL--PREQVVVETKCGI VWERKGS LFNKVGDQLYK---NLS	110
	:	:	:	:
Qy		99	AERVRSKDISELERLDYVDILHCHDIEFGSLDOIVSETIPALQKLQBGKTFRFIGTG	158
Dd		111	PESIREEVEASLQRGLGFDYDIYNTHWSQVPFPYTIAETVAVLNELKAEGKIRAICAA	170
	:	:	:	:
Qy		159	LPLDIFTYVLDRPVGTVDDVILSYCHGVNDSTL-LDLLPYLKSKGVGISASPLAMGLL	217
Dd		171	VDAD--HIREYLQYGELDI--QAQYSILDRA MENELLPLCRDNGIUVQVYFLEQGLL	225
	:	:	:	:
Qy		218	TGOGPPWHHPASPPELKSASKA VAH-----CKSKGKKITKLALQYSLAN	261
Dd		226	TGTTIRDYVPGGAR--ANKVMFORENNLVKMVIDMLEQWPCLCARYOCTIPTTLALAWILKQ	282
	:	:	:	:
Qy		262	KETSSVLVGMSSVSOZEENVAATELESGLMDQTILSEVAIL	304
Dd		283	SDLISLUSGATAPEQVRENVAAL-----NINUSDADATL	316
	:	:	:	:

RESULT 12
D85787
hypothetical protein Z2809 [imported] - Escherichia coli (strain O157:H7, substrain EDL93)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence revision 16-Feb-2001 #text change 09-Jul-2004

C;Accession: D85787
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouxis, K.; Apodaca,
Nature 409, 525-533, 2001
A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A;Reference number: AB5480; MUID:21074935; PMID:11206551
A;Accession: D85787
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-326 <STO>
A;Cross-references: UNIPROT:Q8XDV3; GB:AE005174; NID:g12515802; PIDN:AAG56760.1; GSPDB:(
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z2809
C;Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match	19.4%	Score 314	DB 2	Length 326
Best Local Similarity	30.0%	Pred. No. 4	4e-16	
Matches 103	Conservative	46	Mismatches 128	Indels 66
				Gaps 12

Qy	1	MTKIELRALGNTGLKVSAGVFGSPLGSVFGPVAEDD-----AVATVREAFRLGINFFDT	55
		: : : : : : : : : : : : : : : : : : : :	
Db	1	MKKI--PLGTTDITLSRMGLGTWAGG--GPANNGDLDRQICIDITILEAHRGCLNIDT	55
		: : : : : : : : : : : : : : : : : : : :	
Qy	56	SPYGGTLSRXMLGKGKALQVPRSDYIVATKCG-----RYKSGFDFS	98
		: : : : : : : : : : : : : : : : : : : :	
Db	56	APGYNFCNSEIVGQALKCL--PREQVVVETKCGIWKRGSLPNKVGDRQLYK--NLS	110
		: : : : : : : : : : : : : : : : : : : :	
Qy	99	AERVKSIDSLERLQIDYDILHCHIEFSGLSQIVSETIPALOKLQKQKTRFIGTG	158
		: : : : : : : : : : : : : : : : : : : :	
Db	111	PESTREEVASLQRLGFDYIDIVYTHMQSVPPFTPTAETVAVLNMLNKAEGKIRAIGAN	170
		: : : : : : : : : : : : : : : : : : : :	
Qy	159	LPDIFIVYLDVRPPGTVDTLSYCHYGVDSTL-LDLLPYLKSXGUVGISASPLAMGLL	217
		: : : : : : : : : : : : : : : : : : : :	
Db	171	VDAD--HIREYQYGBELII--QAKYSILDRAMENELLPCRDNGIIVQVVSPLQGLL	225
		: : : : : : : : : : : : : : : : : : : :	
Qy	218	TEQGPSPWHPASPELKSASKAAVAH-----CKSKGKITKILALQVSLAN	261
		: : : : : : : : : : : : : : : : : : : :	
Db	226	TGHTIRDYVGGAR--ANKWVFORENMLKVIDMLEQWQPLCARVQCTIPTILANILKQ	282
		: : : : : : : : : : : : : : : : : : : :	
Qy	262	KEISSLVLGMSVSVQVEENAAVTELSLGMQDTLSEVAIL	304
		: : : : : : : : : : : : : : : : : : : :	
Db	283	SDIISITISGATAPROVENVAAL-----NINLSDADATL	316
		: : : : : : : : : : : : : : : : : : : :	

```

RESULT 13
C64937
hypothetical protein b1771 - Escherichia coli (strain K-12)
C/Species: Escherichia coli
C/Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
C/Accession: C64937
R/Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;
A.; Rose, D.J.; Mau, B.; Shaoh, Y.
Science 277, 1453-1462, 1997
A/Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; MUID:97426617; PMID:9278503
A/Accession: C64937
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-326 <BLAT>
A/Cross-references: UNIPROT:P77256; GB:AE000272; GB:U00096; NID:gl788067; PIDN:AACT4841.
A/Experimental source: strain K-12, substrain MG1655
C/Superfamily: fission yeast pyridoxine 4-dehydrogenase

```

```

Query Match      19.3%; Score 313; DB 2; Length 326;
Best Local Similarity 30.0%; Pred. No. 5.3e-16;
Matches 103; Conservative 46; Mismatches 128; Indels 66; Gaps 12

Qy 1 MTKIELRALGNTGLKVGAVGFGASPLGSGVFGPVAEDD-----AVATVREARFLGNGFPDPT 55
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1 MKKI-----PLGTDTITLRLMGLGTWAIIG--GPAWNGDLDROICTDITILEAHRGCGINLIDT 55

Qy 56 SPYVGGTLSERKMLGKGLKALQVPRSDYIVATKCG-----RYKGGFGDPS 98

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 13, 2005, 05:48:21 ; Search time 96 Seconds
(without alignments)
7735.208 Million cell updates/sec

Title: US-10-606-300-12
Perfect score: 1653
Sequence: 1 atgacgaataagagcttcg.....gtggaatccatcagaactaa 960

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlh
-Q=/cgm2_1/USPTO_spool_h/US10606300/runat_07112005_094818_7295/app_query.fasta_1.1159
-DB=A_Geneseq -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10606300 @CGN 1 1 101 @runat_07112005_094818_7295 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq 16Dec04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1619	97.9	319	AAE11998	AAE11998 Arabidops
2	1619	97.9	319	AAU76343	AAU76343 A. thalia
3	1600	96.8	357	AAG16520	AAG16520 Arabidops
4	1286	77.8	253	AAG16521	AAG16521 Arabidops
5	593.5	35.9	345	ABB67077	ABB67077 Drosophil
6	544.5	32.9	342	ABB71229	ABB71229 Drosophil
7	420.5	25.4	294	ABB60118	ABB60118 Drosophil
8	366	22.1	329	AAK51284	AAK51284 L-fucose
9	339.5	20.5	321	AAK27118	AAK27118 Fucose de
10	337.5	20.4	306	ADS44757	ADS44757 Bacterial

11	329.5	19.9	304	6	ABU17641	Abu17641 Protein e
12	327.5	19.8	325	8	ADN22222	Adn22222 Bacterial
13	322	19.5	308	8	ADN24982	Adn24982 Bacterial
14	322	19.5	329	6	ABU29907	Abu29907 Protein e
15	317	19.2	367	7	ABO77872	AbO77872 Pseudomon
16	316	19.1	307	8	ADS27427	AdS27427 Bacterial
17	316	19.1	346	6	ABU27806	Abu27806 Protein e
18	314	19.0	325	5	ABP65908	ABP65908 Bifidobac
19	313	18.9	326	4	AAU34593	Aau34593 E. coli c
20	313	18.9	326	6	ABU15320	Abu15320 Protein e
21	313	18.9	326	8	ADN18209	Adn18209 Bacterial
22	311	18.8	346	5	AAU49053	Aau49053 Coriolus
23	310.5	18.8	302	6	ABU43680	Abu43680 Protein e
24	310	18.8	311	6	ABU17621	Abu17621 Protein e
25	307.5	18.6	311	8	ADS30346	AdS30346 Bacterial
26	306.5	18.5	332	6	ABU47024	Abu47024 Protein e
27	305.5	18.5	323	4	AAU36253	Aau36253 Pseudomon
28	305.5	18.5	332	6	ABU47649	Abu47649 Protein e
29	302.5	18.3	341	7	ABO81330	AbO81330 Pseudomon
30	302	18.3	305	6	ABU17897	Abu17897 Protein e
31	301.5	18.2	339	8	ADS41735	AdS41735 Bacterial
32	300.5	18.2	320	8	ADS30364	AdS30364 Bacterial
33	296.5	17.9	315	6	ABU30072	Abu30072 Protein e
34	296.5	17.9	332	7	ADC97264	AdC97264 E. faeciu
35	296	17.9	334	6	ADA34792	AdA34792 Acinetoba
36	295	17.8	344	8	ADS23006	AdS23006 Bacterial
37	294.5	17.8	314	8	ADS41580	AdS41580 Bacterial
38	294.5	17.8	342	8	ADS24931	AdS24931 Bacterial
39	294	17.8	333	6	ABU16921	Abu16921 Protein e
40	291.5	17.6	302	6	ABU42693	Abu42693 Protein e
41	291.5	17.6	314	5	ABP38281	ABP38281 Staphyloc
42	291.5	17.6	314	8	ADS04717	AdS04717 Staphyloc
43	291	17.6	328	6	ABU34174	Abu34174 Protein e
44	291	17.6	348	2	AAU52837	Aau52837 Escherich
45	291	17.6	348	4	AAU34474	Aau34474 E. coli c

ALIGNMENTS

RESULT 1
AAE11998
ID : AAE11998 standard; protein; 319 AA.
XX
AC AAE11998;

DT 18-DEC-2001 (first entry)
DE Arabidopsis thaliana L-galactose dehydrogenase (L-galdH) protein.
XX
KW L-galactose dehydrogenase; L-galdH; transgenic organism; ascorbic acid;
herbicide; chromosome 4.
XX
OS Arabidopsis thaliana.
XX
PN WO200172974-A2.
XX
PD 04-OCT-2001.
XX
PF 29-MAR-2001; 2001WO-GB001412.
XX
PR 29-MAR-2000; 2000GB-00007651.
XX
PA (ASCO-) ASCORBEX LTD.
XX
PI Smirnoff N, Wheeler G;
XX
DR WPI; 2001-616482/71.
XX
N-PSDB; AAD19526.
XX
PT Novel L-galactose dehydrogenase protein and nucleic acid sequence
encoding the protein for producing genetically modified plants and
microorganisms with enhanced ability to synthesize ascorbic acid.
XX

PS Claim 1; Page 57-58; 58pp; English.

XX The present invention relates to an isolated protein having L-galactose
 CC dehydrogenase (L-galdH) biological activity. L-galdH nucleic acid is
 CC useful for generating transgenic organisms and modified plants with
 CC enhanced ability to synthesize ascorbic acid. L-galdH facilitate the
 CC production of a plant that has been genetically modified to express a
 CC mutated L-galdH protein which is resistant to herbicides that act against
 CC the naturally occurring L-galdH and to identify and/or design compounds
 CC that are inhibitors of L-galdH. The compounds can be used, for e.g. in a
 CC herbicide which acts on L-galdH and damages or kills plants that express
 CC the enzyme. The present sequence is Arabidopsis thaliana L-galactose
 CC dehydrogenase (L-galdH) protein. The L-galdH gene is located on
 CC chromosome 4

XX SQ Sequence 319 AA;

Alignment Scores:
 Pred. No.: 1.45e-165 Length: 319
 Score: 1619.00 Matches: 319
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 97.94% Indels: 0
 DB: 4 Gaps: 0

US-10-606-300-12 (1-960) x AAE11998 (1-319)

QY 1 ATGACGAAATAGAGCTTCGAGCTTTGGGGAACACAGAGCTTTAAGTTAGCGCCGTTGGT 60
 DB 1 MetThrLysIleGluLeuArgAlaLeuGlyAenThrGlyLeuLysValSerAlaValGly 20
 QY 61 TTTGGTGCTCTCCGCTGGAAAGTGTCTTCGGTCAGTCGCCGGAAGATGATGCCGTCGCC 120
 DB 21 PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAla 40
 QY 121 ACCGCGCGAGGCTTCGCTCGGTATCACTCTTCGCACACCTCCCGTATTATGGA 180
 DB 41 ThrValArgGluAlaPheArgLeuGlyIleAenPheAspThrSerProTyrFyrGly 60
 QY 181 GGACACATGTCTGAGAAATGCTTGGTAAGGAGCTTAAAGGCTTTGCAAGTCCCTAGAGT 240
 DB 61 GlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGluValProArgSer 80
 QY 241 GACTACATTTGGCTACTAAGTGTGGTAGATATAAAGAGGTTTTCATGCTGCTGAG 300
 DB 81 AspTyrIleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGlu 100
 QY 301 AGAGTAAGAAAGATATTGACGAGAGCTTGGAGAGGCTTCAGCTTGATTATGTTGACATA 360
 DB 101 ArgValArgLysSerIleAspGluSerLeuGluArgLeuGluAspTyrValAspIle 120
 QY 361 CTTTCATTCGCATGACATTTGAGTTCGGGTCTCTTGATCAGATTGTGAGTGAAACAATTCCT 420
 DB 121 LeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePro 140
 QY 421 GCTCTTCAGAACTGAACAAGAGGGGAAGACCCGGTTCATTTGATCAGTGTCTTCG 480
 DB 141 AlaLeuGlnLysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGlyLeuPro 160
 QY 481 TTAGATATTTTCACTATGTTCTTCATCGAGTGCTCCAGGACGCTCGATGTGATATTG 540
 DB 161 LeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeu 180
 QY 541 TCATCTGTCTATTCAGGCGCTTAATGATTCGAGTTCGCTGGATTTACTACTTACTTGAAG 600
 DB 181 SerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLys 200
 QY 601 AGCAAGGTGTGGGTGTGATAGTGTCTTCATTAGCATGGGCTCTCTTACAGAACAA 660
 DB 201 SerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGln 220
 QY 661 GGTCTCTCTGTAATGGCACCTCTCTCCCTGAGCTCAAGTCTGCAAGCAAGCCGAGTT 720

Db 221 GlyProProGluTyrHisProAlaSerProGluLeuLysSerAlaSerLysAlaVal 240
 QY 721 GCTCAGTCAAAATCAAGGGCAAGATCAAAAGTTAGCTCTGCAATACAGTTTAGCA 780
 Db 241 AlaHisCysLysSerLysGlyLysIleThrLysLeuAlaLeuGlnTyrSerLeuAla 260
 QY 781 AACAGGAGATTTCGTGGTCTTGGTGGGATGAGCTCTGTCTCACAGTAGAAGAAAT 840
 Db 261 AsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGluAsn 280
 QY 841 GTTCAGCAGATTACAGAGCTTCAAAAGTCTGGGGATGATCAAGAACTCTGTCTGAGTT 900
 Db 281 ValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluVal 300
 QY 901 GAAGCTATTCTCGAGCCCTGTAAAGATCTGACATGGCCCAAGTGAATCCATCAGAAC 957
 Db 301 GluAlaIleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGlnAsn 319

RESULT 2

AAU76343

ID AAU76343 standard; protein; 319 AA.

XX AC AAU76343;

XX DT 21-MAY-2002 (first entry)

XX DE A. thaliana L-galactose dehydrogenase (LGDH).

XX KW Ascorbic acid; vitamin C; scurvy; recombinant yeast; enzyme;

XX KW L-galactose dehydrogenase; LGDH; L-galactono-1,4-lactone dehydrogenase;

XX KW AGD; D-arabinose dehydrogenase; ARA; D-arabinono-1,4-lactone oxidase;

XX KW ALO; L-gulonono-1,4-lactone oxidase; GLO; aldono-lactonase; AL.

XX OS Arabidopsis thaliana.

XX PN WO200210425-A2.

XX PD 07-FEB-2002.

XX PF 02-AUG-2001; 2001WO-GB003485.

XX PR 02-AUG-2000; 2000US-00630983.

XX PA (BIOP-) BIPOLO SCARL.

XX PA (WHAL/) WHALLEY K.

XX PI Porro D, Sauer M;

XX DR WPI; 2002-217125/27.

XX DR N-PSDB; ASK10125.

XX Generating ascorbic acid or its salt, involves culturing yeast capable of
 PT converting ascorbic acid precursor into ascorbic acid in medium
 PT comprising ascorbic acid precursor, and isolating ascorbic acid.
 XX Claim 12; Page 86-87; 95pp; English.

XX The invention relates to generating ascorbic acid or its salt, involves

CC obtaining a recombinant yeast capable of converting an ascorbic acid
 CC precursor into ascorbic acid, culturing the recombinant yeast in a medium
 CC comprising an ascorbic acid precursor, thus forming ascorbic acid, and
 CC isolating the ascorbic acid. Also include are stabilising ascorbic acid
 CC or its salt in a medium, by culturing a yeast in a medium comprising
 CC ascorbic acid or its salt and a recombinant yeast functionally
 CC transformed with a coding region encoding a protein having an enzyme
 CC activity selected from L-galactose dehydrogenase (LGDH), L-galactono-1,4
 CC -lactone dehydrogenase (AGD), D-arabinose dehydrogenase (ARA), D-
 CC arabinono-1,4-lactone oxidase (ALO), L-gulonono-1,4-lactone oxidase (GLO)
 CC or converting to ascorbic acid at least about 25% ascorbic acid precursor
 CC or is capable of producing at least 20 mg ascorbic acid/L medium, when
 CC the yeast is cultured in a medium comprising one ascorbic acid precursor.
 CC The ascorbic acid produced (Vitamin C) is a powerful antioxidant, a

CC deficiency of which causes scurvy in humans. The present sequence
CC represents A. thaliana LGDH

XX
SQ Sequence 319 AA;

Alignment Scores:
Pred. No.: 1,45e-165 Length: 319
Score: 1619.00 Matches: 319
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 97.94% Indels: 0
DB: 5 Gaps: 0

US-10-606-300-12 (1-960) x AAU76343 (1-319)

QY 1 ATGACGAAATAGAGCTTCGAGCTTTGGGGAACACAGGGCTTAAGGTTAGCGCGTTGGT 60
DB 1 MetThrLysLleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGly 20
QY 61 TTGGTGCTCTCCGCTCGGAGTGCTTCGGTCCAGTCGCCGAGATGATGCCGTGCC 120
DB 21 PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAla 40
QY 121 ACCGTGCGGAGGCTTTCCGTCTCGGTATCACTTCGACACCTCCCGTATATGGA 180
DB 41 ThrValArgGluAlaPheArgLeuGlyLysPheAspThrSerProTyrTyrGly 60
QY 181 GGACACATGCTCGAGAAATGCTTGAAGGACTTAAAGGCTTTCGAAAGTCCCTAGAGT 240
DB 61 GlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSer 80
QY 241 GACTACATTGTGGCTACTAAGTGTGGTATATATAAGAGGTTTGTATTTTCAGTCTGAG 300
DB 81 AspTyrLleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGlu 100
QY 301 AGAGTAAGAAGATGATTACAGAGCTTCGAGAGCTTCAGCTTGATGATATGTCACATA 360
DB 101 ArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 120
QY 361 CTTCATTCGCCATGACATTGAGTTCGGTCTCTTGATCAGATTGTGAGTGAAACAATTCCT 420
DB 121 LeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePro 140
QY 421 GCTCTTCAGAACTGAACAAGAGGGGAGACCCGGTTTCATTGGTATCACTGGCTTCG 480
DB 141 AlaLeuGlnLysLeuLysGlnGluGlyLysThrArgPheLleGlyLleThrGlyLeuPro 160
QY 481 TTAGATATTTCACTTATGTTCTTGATCGAGTCCCTCCAGGACGTGTCGATGTGATATG 540
DB 161 LeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeu 180
QY 541 TCATATCTCTATTACGGGTAAATGATTCGAGTTCGGATTCTACTACCTTACTTGAAG 600
DB 181 SerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLys 200
QY 601 AGCAAGGTGTGGTGTGATAGTCTCTCCATTAGCAATGGGCTCTCTTACAGAACAA 660
DB 201 SerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGln 220
QY 661 GGTCCTCTGAATGACACCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAGCCGCGATT 720
DB 221 GlyProProGluTrpHisProAlaSerProGluLeuLysSerAlaSerLysAlaVal 240
QY 721 GCTCACTCCAATCAAGGCAAGAGATCAAAAGTTAGCTCTCAATACAGTTTACGA 780
DB 241 AlaHisCysLysSerLysGlyLysLysIleThrLysLeuAlaLeuGlnTyrSerLeuAla 260
QY 781 AACAGGAGATTTCCGTGTTGGTGGATGAGCTCTGTCTCACAGGTAGAGAAAT 840
DB 261 AsnLysGluLleSerSerValLeuValGlyMetSerSerValSerGlnValGluGlnAsn 280
QY 841 GTTCAGAGTTACAGAGCTTGAAGTCTGGGATGGATCAAGAACTCTGTCTGAGTT 900

Db 281 ValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluVal 300
QY 901 GAAGCTATTTCTCGAGCTCTGTAAGAATCTGACATGGCCCAAGTGAATCCATCAGAAC 957
DB 301 GluAlaIleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGlnAsn 319
RESULT 3
AAG16520
ID AAG16520 standard; protein; 357 AA.
XX
AC AAG16520;
XX
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 17197.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS
PN EP1033405-A2.
PD 06-SEP-2000.
PF 25-FEB-2000; 2000EP-00301439.
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.

PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142300P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.

PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:

Pred. No.: 1.74e-163
Score: 1600.00
Length: 357
Matches: 317
Percent Similarity: 99.69%
Conservative: 1
Best Local Similarity: 99.37%
Mismatches: 0
Query Match: 96.79%
Indels: 1
Gaps: 0
DB:

US-10-606-300-12 (1-960) x AAG16520 (1-357)

QY	3	GACGAAATAGAGCTTCGAGCTTTCGGGGAACAAGGCGCTTAAGGTTAGCGCGCTTGCTTT	62
DB	40	AspGluAsnArgAlaSer-AlaLeuGluGlyAsnThrGlyLeuLysValThrAlaValGly	59
QY	63	TGSGTCTCTCCGCTCGCAAGTGTCTTCGGTCCAGTCGCCGAAGATGATGCGCTGCGCAC	122
DB	59	edgylalaserProLeuGlySerValPheGlyProValaladgluaspaspalavalalath	79
QY	123	CGTGGCGAGGCTTTCCGTCTCTCGGTATCAACTCTTCGACACCTCCCGCTATTATGAGG	182
DB	79	rValArgGluAlaPheArgLeuGlyLleAenPheAspThrSerProTyrTrgIygl	99
QY	183	AACACTGTCTGAGAAATCTTGTTAAGGAGCTTAACGCTTCGACAGTCCCTAGAGTGA	242
DB	99	YthrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAs	119
QY	243	CTACATTGCGCTACTAAGTGTGTGTAGATATAAAGAAGTTTGTATTTCAGTGTCTGAG	302
DB	119	pyrillevalAlathrlsCyGlyArgTyrLysgluglyPheAspPheSerAlaGluar	139
QY	303	AGTAAGAAGAGATTATGACGAGAGCTTCGGAGAGCTTCAGCTTCGATTATGTCACATCT	362
DB	139	gValarglysSerlleaspgLuserLeuGluArgLeuGlnLeuAspTyrValaspilele	159
QY	363	TCATTGCCATGACATTGATTGCGGTCTCTTGATCAGATTGTGAGTGAAACAATTCCTGC	422
DB	159	uhisCyshisAspilleGluPheGlySerLeuAspGlnIleValSerGluThrIleProAl	179
QY	423	TCCTTCAGAACTGAACAGAGGGGAGACCCGGTTCATTGGTATCACTGGCTTCGGTT	482
DB	179	aleuGlnLysLeuLysGlnGluGlyLysThrArgPheilleGlyIleThrGlyLeuProle	199
QY	483	AGATATTTTCACTATGTTCTTGATTCGAGTGCCTCCAGGACTGTGCGTGTGATTTGTC	542
DB	199	uaspilePheThrYrValLeuAspArgValProProGlyThrValAspValilleuSe	219
QY	543	ATACTGTCATTACGGCGTTAATGATTCGAGTCTCTCGATTTACTACCTTACTTGAAG	602
DB	219	rTyrCyshisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLysSe	239
QY	603	CAAAGTGTGGTGTGATAAGTGTCTTCCATTAGCAATGGGCTCTCTACAGAACAGG	662
DB	239	rLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGlnGl	259
QY	663	TCCTCTCAATGGCACCTGCTTCCTCTGAGCTCAAGTCTGCAAGCAAAGCCGACGTGC	722
DB	259	yProProGluTrpHisProAlaSerProGluLeuLysSerAlaSerlyseAlaValal	279
QY	723	TCATCGCAANTCAAGGGCAAGAAGATCACAAAGTTAGCTCTGCAATACAGATTTAGCAA	782
DB	279	ahisCyshysSerLysGlyLysLysIleThrLysLeuAlaLeuGlnTyrSerLeuAlaAs	299
QY	783	CAAGAGATTCTCGTCGGTGTGGTGGATGAGCTCTGTCTCAGGTAGAGAAATGT	842
DB	299	nlYsGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGluAsnVa	319
QY	843	TGCACAGTTACAGAGCTTGAAGTCTGGGAGTCGATCAAGAACCTCTGTCTGAGGTTCA	902
DB	319	lAlaAlaValThrGluLeuGluSerLeuGlyMeCaspGlnGluThrLeuSerGluValGl	339
QY	903	AGCTATTCTTCGAGCTGTAAAGAATCTGACATGCGCAAGTGGAAATCCATCAGAAC	957
DB	339	uAlaIleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGlnAsn	357
RESULT 4			
AAG16521			
XX	ID	AAG16521	standard; protein; 253 AA.
XX	AC	AAG16521;	
XX	AC	AAG16521;	
DT	DT	17-OCT-2000	(first entry)
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 17198.		

PR	23-JUN-1999;	99US-0140353P.
PR	23-JUN-1999;	99US-0140354P.
PR	24-JUN-1999;	99US-0140695P.
PR	28-JUN-1999;	99US-0140823P.
PR	29-JUN-1999;	99US-0140991P.
PR	30-JUN-1999;	99US-0141287P.
PR	01-JUL-1999;	99US-0141842P.
PR	01-JUL-1999;	99US-0142154P.
PR	02-JUL-1999;	99US-0142055P.
PR	06-JUL-1999;	99US-0142390P.
PR	08-JUL-1999;	99US-0142803P.
PR	09-JUL-1999;	99US-0142920P.
PR	12-JUL-1999;	99US-0142977P.
PR	13-JUL-1999;	99US-0143542P.
PR	14-JUL-1999;	99US-0143624P.
PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144085P.
PR	16-JUL-1999;	99US-0144086P.
PR	19-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	22-JUL-1999;	99US-0145192P.
PR	23-JUL-1999;	99US-0145145P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145222P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	27-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148315P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.
Alignment Scores:		
Pred. No.:	1.38e-129	Length: 253
Score:	1286.00	Matches: 253
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	77.80%	Indels: 0
DB:	3	Gaps: 0
US-10-606-300-12 (1-960) x AAG16521 (1-253)		
QY	199	ATGCTTGTAAGGAGCTAAAGCGCTTTGCAAGTCCCTAGAAGTGACTACATTGTGGCTACT 258
Db	1	MetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAspTyrIleValAlaThr 20
QY	259	AAGTGTGTAGATATAAAGAAGGTTTTCAGTTTCAGTGTGAGAGAGTAAAGAAGATATT 318
Db	21	LysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGluArgValArgLysSerIle 40
QY	319	GACGAGAGCTTGGAGAGCGCTTCAGCTTCATTATGTTGACATCTTCATTGCCATGACATT 378
Db	41	AspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIleLeuHisCysHisAspIle 60
QY	379	GAGTTCCGGGTCTCTTTGATCAGATTGTGAGTGAACAATTCCTGCTCTTTCAGAAACTGAAA 438
Db	61	GlupheGlySerLeuAspGlnIleValSerGluThrIleProAlaLeuGlnLysLeuLys 80

QY 439 CAAGAGGGAGACCCGGTTTCATGGTATCACTGCTCTCCGTTAGATATTTTCACTTAT 498
 Db 81 GlnGluGlyLeuThrArgPheilegylleThrGlyLeuProLeuAspPheThrTyr 100
 QY 499 GTTCTTGATGAGTCCCTCCAGGACTCCGATGATATGTCATATGTCATTCACGGC 558
 Db 101 ValLeuAspA-gValProProGlyThrValAspValIleLeuSerTyrCysHisTyrGly 120
 QY 559 GTTAATGATTCGACCTTCGCTGCTGATTTACTACTTCAAGAGCAAGAGGTGGGTG 618
 Db 121 ValAsnAspSerThrLeuLeuAspLeuProTyrLeuLysSerLysGlyValGlyVal 140
 QY 619 ATAAGTCTTCTCCATTAGCAATAGGCTCTCTTACAGAACAGGTCTCTGATGGCAC 678
 Db 141 IleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGlnGlyProProGluTrpHis 160
 QY 679 CCTGCTCCCTGAGTCAAGTCTGCAAGCAAGCGCAGTGTCTCACTGCAATCAAG 738
 Db 161 ProAlaSerProGluLeuLysSerAlaSerLysAlaAlaValAlaHisCysLysSerLys 180
 QY 739 GCGAAGAGATCACAAAGTTAGCTCTGCAATACAGTTTAGCAAAACAGGAGATTTCGTCG 798
 Db 181 GlyLysLysIleThrLysLeuAlaLeuGlnTyrSerLeuAlaAsnLysGluLysSer 200
 QY 799 GTGTTGGTGGATGAGTCTGCTCAAGTAGAAGAAATGTTGCAAGTGTACAGAC 858
 Db 201 ValLeuValGlyMetSerSerValSerGlnValGluGlnAsnValAlaAlaValThrGlu 220
 QY 859 CTTGAAGTCTGGGATGATCAAGAACTCTGCTGAGGTTGAAGCTATTCTCGAGCCT 918
 Db 221 LeuGluSerLeuGlyMetAspGlnGlnThrLeuSerGluValGluAlaIleLeuGluPro 240
 QY 919 GTAAAGATCTGACATGCGCAAGTGGATGGAATCCATCAGAAC 957
 Db 241 ValLysAsnLeuThrTrpProSerGlyIleHisGlnAsn 253

RESULT 5
 ABB67077
 ID ABB67077 standard; protein; 345 AA.
 XX AC ABB67077;
 XX DT 26-MAR-2002 (first entry)
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 28023.
 XX KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
 XX OS Drosophila melanogaster.
 XX PN WO200171042-A2.
 XX PD 27-SEP-2001.
 XX PF 23-MAR-2001; 2001WO-US009231.
 XX PR 23-MAR-2000; 2000US-0191637P.
 XX PR 11-JUL-2000; 2000US-00614150.
 XX PA (PEKE) PE CORP NY.
 XX PI Venter JC, Adams M, Li PWD, Myers EW;
 XX DR WPI, 2001-656860/75.
 XX DR N-P5DB; ABL11180.
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.
 XX PS Disclosure; SEQ ID NO 28023; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA sequences (ABU01840-ABU16175) and the encoded proteins (ABB57737-ABB72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 345 AA;
 SQ

Alignment Scores:
 Pred. No.: 1.32e-54 Length: 345
 Score: 593.50 Matches: 128
 Percent Similarity: 63.46% Conservative: 63
 Best Local Similarity: 42.52% Mismatches: 99
 Query Match: 35.90% Indels: 11
 DB: 4 Gaps: 7

US-10-606-300-12 (1-960) x ABB67077 (1-345)

QY 7 AAATAGAGCTTCGAGCTTTGGGAAACACAGGGCTTAAGTTAGCGCGTGGTTTGGT 66
 Db 21 ArgMetGluTyrArgAsnLeuGlyLysThrGlyLeuGlnValSerLysValSerPheGly 40
 QY 67 GCCTCTCGCTCGGAAGTCTTCGCTCCAGTCGCCGAAGATGATCGCTCGCCACCGTG 126
 Db 41 GlyGlyAlaLeuCysAlaAsnTyrGlyPheAspLeuGlu---GluGlyLeuThrVal 59
 QY 127 CGCAGGCTTTCCTCGGTATCAACTTCTTCGACACCTCCCGGTATTTAGGAGGAACA 186
 Db 60 HisGluAlaValLysSerGlyIleAsnTyrIleAspThrAlaProTyrTyrGlyGlnGly 79
 QY 187 CTGCTCAGAAATGCTTGGTAAGGACTAAAGCTTTGCAAGTCCCTAGAGTGACTAC 246
 Db 80 ArgSerGluGluValLeuGlyLeuAlaLeuLys-----AspValProArgGluSerTyr 97
 QY 247 ATTGTGGCTACTAAGTGTGTAGATATAA-----GAAGGTTTTGATTTCAGT 294
 Db 98 TyrIleAlaThrLysValAlaArgTyrGluLeuAspTyrAspLysMetPheAspPheSer 117
 QY 295 GCTGAGAGAGTAAGAAAGAGTATTGACGAGAGCTTTGGAGAGGCTTCAGCTTTGATTAT 354
 Db 118 AlalysLysThrArgGluSerValGluLysSerLeuLysLeuLeuGlyLeuAspTyrVal 137
 QY 355 GACATATCTATTCCTCCATGACATTGAGTTCGGG---TCTCTTGATCAGATTGTGAGTCAA 411
 Db 138 AspValIleGlnIleHisAspIleGluPheAlaLysAspLeuAspIleValIleAsnGlu 157
 QY 412 ACAATTCCTGCTCTTCAGAACTCAAAACAGAGGGAGACCCGGTTCATTGGTATCACT 471
 Db 158 ThrLeuProThrLeuGluGlnLeuValGlyGluGlyLysAlaArgPheIleGlyValSer 177
 QY 472 GGTCTTCGTTAGATATTTTCACCTTATGTTCTTCGATGAGTGCCTCCAGGAGCTGTGAT 531
 Db 178 AlaTyrProIleSerValLeuLysGluPheLeuThrArgThr---AlaGlyArgLeuAsp 196
 QY 532 GTGATATTGTCACTATCTGCTATTCAGCGGTAAATGATTCGAGCTGTGGATTACTACT 591
 Db 197 ThrValLeuThrTyrAlaArgTyrThrLeuThrAspGluThrLeuLeuGluTyrLeuAsp 216
 QY 592 TACTTTGAAGCAAGGTGTGGGTGATGATGCTCTTCTCCATTAGCAATGGGCTCCTT 651
 Db 217 PhePheLysSerGlnAsnLeuGlyValIleCysAlaAlaAlaHisAlaLeuGlyLeuLeu 236
 QY 652 ACAGAACAAAGTCTCTGTAATGCAACCTGCTCTCCCTGAGCTCAAGTCTGCAAGCAAA 711
 Db 237 ThrAsnAlaGlyProGlnProTrpHisProAlaSerAspGluGlnLysAlaIleAlaArg 256
 QY 712 GCCCGAGTTGCTCTACTGCAANTCAAGGGCAAGAGATCACAAGATTAGTCTCTGCAATAC 771

```

Db      257 LysAlaSerGluValCysLeuGluArgGlyValGluLeuGlyLysLeuAlaMetTyrTyr 276
QY      772 AGTTTAGCAAAAC---AAGGAGATTTCGTCGGTGTGGTGGGATGAGCTCTGTCTCACAG 828
Db      277 ThrMetSerGlyLeuProGluValSerThrPheLeuThrGlyMetGlnThrArgGlnLeu 296
QY      829 GTAGAGAAATGTTGACAGCTTACAGAGCTTGAAGTCTGGGATGATGATCAAGAAACT 888
Db      297 LeuArgIleAsnLeuAspAla---AsnGluValGlyLeuSerAspLysGluGlnGluVal 315
QY      889 CTG 891
Db      316 Leu 316

RESULT 6
AB871229
ID AB871229 standard; protein; 342 AA.
XX
AC AB871229;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 40479.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
WPI; 2001-656860/75.
DR N-PSDB; ABL15332.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 40479; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 342 AA;

Alignment Scores:
Pred. No.: 2.63e-49 Length: 342
Score: 544.50 Matches: 117
Percent Similarity: 58.75% Conservative: 117
Best Local Similarity: 36.56% Mismatches: 117
Query Match: 32.94% Indels: 15
DB: 4 Gaps: 7

New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signaling and cell-cell
interactions.
Disclosure; SEQ ID NO 40479; 21pp + Sequence Listing; English.
The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
ABB72072). The sequence data for this patent did not form part of the
printed specification, but was obtained in electronic format directly
from WIPO at ftp.wipo.int/pub/published_pct_sequences
Sequence 342 AA;

Alignment Scores:
Pred. No.: 2.63e-49 Length: 342
Score: 544.50 Matches: 117
Percent Similarity: 58.75% Conservative: 117
Best Local Similarity: 36.56% Mismatches: 117
Query Match: 32.94% Indels: 15
DB: 4 Gaps: 7

```

```

US-10-606-300-12 (1-960) x ABB71229 (1-342)
QY      7 AAAATAGAGCTTCGAGCTTTGGGAAACACAGGCTTAAGTTAGCGCTGGTTGGTTGGT 66
Db      21 ArgMetGluTyrArgGlnLeuGlySerThrGlyLeuArgValSerLysIleAlaLeuGly 40
QY      67 GCCTCTCGCTCGGAAGTGTCTTCGGT---CCAGTCCCGAAGATGATGCGCTCGCCACC 123
Db      41 GlyAlaThrLeuSerLysLeuPheSerAspPheAspArgGluGluGlyLeuLeuThr 60
QY      124 GTGCGCAGGCTTCCGCTCGGTATCACTTCTCGACACTCTCCCTCTATTATATGAGGA 183
Db      61 ValGlnGluAlaIleArgSerGlyIleAsnThrIleAspThrAlaProPheTyrGlyGln 80
QY      184 ACACCTGTCTGAGAAAATGCTTGGTAAGGGACTAAAGGCTTTGCAAGTCCCTAGAGTAGC 243
Db      81 GlyLysSerGluGluLeuLeuGlyGlnAlaLeuLys-----AspValProArgGluAla 98
QY      244 TACATTGTGGCTACTAAGTGTGGTAGATATAA-----GAAGGTTTGTGATTTC 291
Db      99 TyrTyrIleAlaThrLysValAlaArgTyrGluLeuAspProAsnAsnMetPheAspTyr 118
QY      292 AGTCTGAGAGAGTAAGAAAGAGTATTGACGAGAGCTTGGAGAGCTTCAGCTTCATTAT 351
Db      119 ThrAlaAlaLysAlaArgGluSerValLysArgSerLeuGluLeuGlnLeuAspArg 138
QY      352 GTTGACATPACTTCATTCGACATTCAGATTGAGTTCGG---TCTCTTGATCAGATTGTGAT 408
Db      139 ValAspValLeuGlnValHisAspValAspAlaAlaProSerLeuAspMetValLeuAsn 158
QY      409 GAACAATTCCTGCTCTTCAGAAACTGAAACAGAGGGGAGAGACCGGTTTCATTGGTATC 468
Db      159 GluThrIleProValLeuGluGluTyrValGlnAlaGlyLysAlaArgPheIleGlyVal 178
QY      469 ACTGCTCTTCGGTTAGATATTTCATTATGTTGATCGAGTCCCTCCAGGGAGCTGTC 528
Db      179 ThrAlaTyrAspValAspValLeuLysGluCysAlaGluArg--GlyLysGlyArgIle 197
QY      529 GATGTGATATTGTTCATCTGTCATTACGGCGTTTAATGATTTCGAGCTTCTGTGATTTACTA 588
Db      198 GlnValValLeuAsnTyrAlaArgTyrThrLeuLeuAspAsnThrLeuLeuArgHisMet 217
QY      589 CTTTACTTGAAGACCAAGGTGGTGGTGTGATAGTGTCTCTCCATTAGCAATGGGCTC 648
Db      218 LysAlaPheGlnGluMetGlyValGlyValCysAlaAlaAlaHisSerLeuGlyLeu 237
QY      649 CTTACAGAAACAAGGTCTCTGAAATGGCACCTGCTTCCCTCGAGCTCAAGTCTCAAGC 708
Db      238 LeuSerAsnAlaGlyProGlnSerTrpHisProGlySerProGluLeuLeuAlaValGly 257
QY      709 AAAGCCGAGTTGCTCACTGCAAAATCAAAAGGGCAAGAGATCACAAAGTTAGCTCTGCA 768
Db      258 LysArgGlyAlaGluIleCysGlnLysArgAsnValGluLeuGlyLysLeuAlaMetTyr 277
QY      769 TACAGTTTAGCAACAAGAGATTTCGTCGGTGTGGTGGATGAGCTCTGTCTCACAG 828
Db      278 TyrThrMetGlnLeuAspGlyAlaAlaThrPheLeuIleGlyIleProAsnArgLysLeu 297
QY      829 GTAGAAGAAAATGTTGACAGCTTACAGAGCTTGAAGCTCTGGGATGATGATCAAGAACT 888
Db      298 LeuArgIleAsnLeuAspAlaIlePheAsp-----GlyLeuThrSerHisGlu 313
QY      889 CTGCTGAGGTTGAAGCTATTCTCGAGCTGTGA-----AGAATCTGACATGGCCAGT 942
Db      314 GlnGluValLeuGlnTyrLeuArgGluAsnValPheThrLysSerTyrSerTrpGlySer 333

RESULT 7
AB860118
ID AB860118 standard; protein; 294 AA.
XX
AC AB860118;
XX
DT 26-MAR-2002 (first entry)

```


QY	22	GCTTTGGGGAA	CACAGGGCTTAAAGGTTAGCCCGCTGGTTTGGTGGCTCTCCGCTCGGA	81
Db	7	AlaAlaAlaAlaAlaGlyLeuAlaIleProAlaLeuGlyTyrGlyAlaAlaAsnValGly	7	AlaAlaAlaAlaAlaGlyLeuAlaIleProAlaLeuGlyTyrGlyAlaAlaAsnValGly
QY	82	AGTGTCTTCGGTCCAGTCGCGAGATGATCCGTCGCCACCGTCGCGAGGCTTCCGT	141	AGTGTCTTCGGTCCAGTCGCGAGATGATCCGTCGCCACCGTCGCGAGGCTTCCGT
Db	27	AsnLeuPheArgAlaLeuSerAspGluAlaTrpAlaValLeuGluAlaAlaTrpAsp	46	AsnLeuPheArgAlaLeuSerAspGluAlaTrpAlaValLeuGluAlaAlaTrpAsp
QY	142	CTCGGTATCAACTTCTTCACACTCCCGCTATTATTCGAGGAACTGCTCGAAAAATG	201	CTCGGTATCAACTTCTTCACACTCCCGCTATTATTCGAGGAACTGCTCGAAAAATG
Db	47	AlaGlyIleArgTyrTyrAspThrAlaProHisTyrGlyLeuGlyLeuSerGluLysArg	66	AlaGlyIleArgTyrTyrAspThrAlaProHisTyrGlyLeuGlyLeuSerGluLysArg
QY	202	CTTGGTAAGGGACTAAAGCTTTGCACCTCCCTAGAGTGCATACATTGTGGCTACTAAG	261	CTTGGTAAGGGACTAAAGCTTTGCACCTCCCTAGAGTGCATACATTGTGGCTACTAAG
Db	67	LeuGlyAlaPheLeuGlnThr-----LysProArgAspGluPheValValSerThrLys	84	LeuGlyAlaPheLeuGlnThr-----LysProArgAspGluPheValValSerThrLys
QY	262	TGTGGTAGA-----	270	TGTGGTAGA-----
Db	85	AlaGlyArgLeuLeuArgProAsnProGluArgArgProSerGlyLeuAspThrAspAsn	104	AlaGlyArgLeuLeuArgProAsnProGluArgArgProSerGlyLeuAspThrAspAsn
QY	271	-----TATAAAGAAGGTTTTGATTTTCAGTCGTGAGAGAGTAAGA	309	-----TATAAAGAAGGTTTTGATTTTCAGTCGTGAGAGAGTAAGA
Db	105	AspPheHisValProAspAspLeuArgArgGluTrpAspPheThrGluGlnGlyIleArg	124	AspPheHisValProAspAspLeuArgArgGluTrpAspPheThrGluGlnGlyIleArg
QY	310	AGAAGTATTACAGAGCTTGAGAGGCTTCAGCTTGATTGTTGTGACATCTCATTCG	369	AGAAGTATTACAGAGCTTGAGAGGCTTCAGCTTGATTGTTGTGACATCTCATTCG
Db	125	AlaSerIleAlaGluSerGlnGluArgLeuGlyLeuAspArgIleAspLeuLeuTyrLeu	144	AlaSerIleAlaGluSerGlnGluArgLeuGlyLeuAspArgIleAspLeuLeuTyrLeu
QY	370	CATGACATTGAGTTCGGGTCTCTTCATCAGATTGTGAGTGAAACAAATCTCGCTCTTCAG	429	CATGACATTGAGTTCGGGTCTCTTCATCAGATTGTGAGTGAAACAAATCTCGCTCTTCAG
Db	145	HisAspProGluArgHisAspLeuAspLeuAlaLeuAlaSerAlaPheProAlaLeuGlu	164	HisAspProGluArgHisAspLeuAspLeuAlaLeuAlaSerAlaPheProAlaLeuGlu
QY	430	AAACTGAAACAGAGGGGAAGACCCGGTTCAATTGGTATCATCTGCTTCCTCGTTAGATATT	489	AAACTGAAACAGAGGGGAAGACCCGGTTCAATTGGTATCATCTGCTTCCTCGTTAGATATT
Db	165	LysValArgAlaGluGlyValValLysAlaIleGlyIleGlySerMetValSerAspAla	184	LysValArgAlaGluGlyValValLysAlaIleGlyIleGlySerMetValSerAspAla
QY	490	TTCACTTATGTTTGATCGAGTGCTCCAGGAGCTCGATGTGATATTGTCATCTACTGT	549	TTCACTTATGTTTGATCGAGTGCTCCAGGAGCTCGATGTGATATTGTCATCTACTGT
Db	185	LeuThrArgAla-----ValArgGluAlaAspLeuAspLeuIleMetValAlaGly	201	LeuThrArgAla-----ValArgGluAlaAspLeuAspLeuIleMetValAlaGly
QY	550	CATTACGGCGTTAATGATTCGACGTTGCTG-----GATTTACTACCT	591	CATTACGGCGTTAATGATTCGACGTTGCTG-----GATTTACTACCT
Db	202	ArgTyr-----ThrLeuLeuGluGlnProAlaAlaThrGluValLeuPro	216	ArgTyr-----ThrLeuLeuGluGlnProAlaAlaThrGluValLeuPro
QY	592	TACTTGAAGAGCAAGGTGTGGGTGTATAAGTCTTCTCCATTAGCAATGGGCGCTCCTT	651	TACTTGAAGAGCAAGGTGTGGGTGTATAAGTCTTCTCCATTAGCAATGGGCGCTCCTT
Db	217	AlaCysAlaGluAsnAlaThrGlyIleValAlaAlaSerValPheAsnSerGlyLeuLeu	236	AlaCysAlaGluAsnAlaThrGlyIleValAlaAlaSerValPheAsnSerGlyLeuLeu
QY	652	ACGAACAAGGTCCTCCT-----GAATGGCACCTCGCTTCCCTCGAGCTC	696	ACGAACAAGGTCCTCCT-----GAATGGCACCTCGCTTCCCTCGAGCTC
Db	237	AlaGlnSerGluProLysArgAspGlyArgTyrGluTyrGlyGlnLeuProAspGluLeu	256	AlaGlnSerGluProLysArgAspGlyArgTyrGluTyrGlyGlnLeuProAspGluLeu
QY	697	AGTCTGCAAGCAAAAGCCGAGTGTGCTCACTGCAAAATCAAGGGCAAGAGATCACAAG	756	AGTCTGCAAGCAAAAGCCGAGTGTGCTCACTGCAAAATCAAGGGCAAGAGATCACAAG

[illegible]


```
US-10-606-300-12 (1-960) x ADN24982 (1-308)
Qy 49 AGCCCGTGTGTTTGGTGGCTCTCCGCTCGAAGTGTCTTCAGTCCGCGCAAGAT 108
Db 1 ThrValMetGlyPheGlyAlaAlaProIleGlyAsnIlePheArgProValSerGluGlu 20
Qy 109 GATCCGTCGCGCCGCGCGAGGCTTTCGCTCGGTATCAACTTCTTCGACACCTCC 168
Db 21 AspSerAlaAlaLeuIleLysAlaAlaTrpAspAlaGlyValArgTyrPheAspThrAla 40
Qy 169 CCGTATTATGAGCAACACTCTCTCAGAAATGCTTGTAGGCACTAAAGCTTTGCAA 228
Db 41 ProMetTyrGlyHisGlyLeuSerGluValArgCysGlyGlnGlyLeuArg-----Trp 58
Qy 229 GTCCCTAGAGTACTACATTGTGCTACTAAGTGTGTAG----- 270
Db 59 TyrProArgAspGlnTyrValLeuSerThrLysValGlyArgLeuLeuLysProArgArg 78
Qy 271 -----TATAAAGAAAGGTTTGTAT 288
Db 79 ArgAlaGluIleAspPheAlaProTrpValAspGlyLeuProPheGluProValPheAsp 98
Qy 289 TTCAGTCTGACAGATGAGAAAGTATTACAGAGCTTGGAGAGCTTCACCTTCAT 348
Db 99 TyrSerTyrAspGlyThrMetArgSerIleGluAspSerLeuGlnArgLeuAlaLeuGlu 118
Qy 349 TATGTGATCACTTTCATTCGTCATGACATTCAGTTCGGGTCT----- 390
Db 119 HisIleAspIleAlaLeuIleHisAspIleAspValPheThrHisGlyGluArgGlnPro 138
Qy 391 -----TTTGATCAGATTGTGAGCAACAATTCCTGCTCTTCAGAACTGAAACAAG 444
Db 139 GluMetPheGluAlaAlaMetAlaGlyAlaSerLysAlaLeuLeuLysLeuArgAspGlu 158
Qy 445 GCGAAGACCCGGTTCATTCGTGATCACTGCTTCGTTGATATATTTTCATTTGTTCTT 504
Db 159 GlyValValLysAlaVal-----GlyLeuGlyValAsnGluTrpGlnValAlaHis 175
Qy 505 GATCAGTGCCTCCAGGACTGTCATGTGATATTTGTCATCTGTCATTACGGGTTAAT 564
Db 176 GluAlaIleArgArgGlnAspPheAspCysLeuLeuLeuAlaGlyArgTyrThrLeuLeu 195
Qy 565 GATTGCGAGTTGCTGGAT---TTACTACTTACTTGAAGACAAAGGTGGTGTGATA 621
Db 196 GluGlnAspAlaLeuAspGlyPheLeuProLeuCysGlyLysLysGlnValSerValIle 215
Qy 622 AGTCTTCTCATTAGCAATGGCTCTTACAGAACAGTCTCTCT----- 669
Db 216 LeuGlyGlyGlyTyrAsnSerGlyIleLeuAlaThrGlyAlaValProGlyAlaLysTyr 235
Qy 670 GAATGGCACCTGCTTCCCTGAGTCAAGTCTGCAAGCAAGCGCAGTTGCTCACTGC 729
Db 236 AsnTyrAlaProAlaProGluAlaIleLeuGluArgValArgLysMetGluGlnValCys 255
Qy 730 AAATCAAGGCAAGAGATCACAAGTTAGTCTGCAATACAGTTTACGACAAAGGAG 789
Db 256 ArgGluPheSerValProLeuLysAlaAlaSerLeuGlnPheValLeuGlyHisProAla 275
Qy 790 ATTTCGTCGGTGGTGGATGAGCTCTCTCAGAGTGGTAGAAGAAATGTT 843
Db 276 IleProThrAsnIleProGlyValArgThrValAlaGlnLeuGluAspAsnLeu 293

RESULT 14
ABU29907
ID ABU29907 standard; protein; 329 AA.
XX
AC ABU29907;
XX
XX Best Local Similarity: 31.14%
XX Query Match: 19.48%
XX DB: 6
XX US-10-606-300-12 (1-960) x ABU29907 (1-329)

Enterococcus faecium.
W0200277183-A2.
03-OCT-2002.
21-MAR-2002; 2002WO-US009107.
21-MAR-2001; 2001US-00815242.
06-SEP-2001; 2001US-00948993.
25-OCT-2001; 2001US-0342923P.
08-FEB-2002; 2002US-00072851.
06-MAR-2002; 2002US-0362699P.
(ELIT-) ELITRA PHARM INC.
Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
WPI; 2003-029926/02.
N-PSDB; ACA33777.
New antisense nucleic acids, useful for identifying proteins or screening
for homologous nucleic acids required for cellular proliferation to
isolate candidate molecules for rational drug discovery programs.
Claim 25; SEQ ID NO 57831; 1766pp; English.
The invention relates to an isolated nucleic acid comprising any one of
the 6213 antisense sequences given in the specification where expression
of the nucleic acid inhibits proliferation of a cell. Also included are:
(1) a vector comprising a promoter operably linked to the nucleic acid
encoding a polypeptide whose expression is inhibited by the antisense
nucleic acid; (2) a host cell containing the vector; (3) an isolated
polypeptide or its fragment whose expression is inhibited by the
antisense nucleic acid; (4) an antibody capable of specifically binding
the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
proliferation or the activity of a gene in an operon required for
proliferation; (7) identifying a compound that influences the activity of
the gene product or that has an activity against a biological pathway (8)
required for proliferation, or that inhibits cellular proliferation; (8)
identifying a gene required for cellular proliferation or the biological
pathway in which a proliferation-required gene or its gene product lies
or a gene on which the test compound that inhibits proliferation of an
organism acts; (9) manufacturing an antibiotic; (10) profiling a
compound's activity; (11) a culture comprising strains in which the gene
product is overexpressed or underexpressed; (12) determining the extent
to which each of the strains is present in a culture or collection of
strains; or (13) identifying the target of a compound that inhibits the
proliferation of an organism. The antisense nucleic acids are useful for
identifying proteins or screening for homologous nucleic acids required
for cellular proliferation to isolate candidate molecules for rational
drug discovery programs, or for screening homologous nucleic acids
required for proliferation in cells other than S. aureus, S. typhimurium,
K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
the target prokaryotic essential genes. Note: The sequence data for this
patent did not form part of the printed specification, but was obtained
in electronic format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences
Sequence 329 AA;
Alignment Scores:
Pred. No.: 3,05e-25 Length: 329
Score: 322.00 Matches: 104
Percent Similarity: 49.10% Conservative: 60
Best Local Similarity: 31.14% Mismatches: 124
Query Match: 19.48% Indels: 46
DB: 6 Gaps: 12
US-10-606-300-12 (1-960) x ABU29907 (1-329)
```

```
QY 7 AAATAGAGCTTCGAGCTTTGGGAAACACAGGCTTAAGTTAGCGCCGTTGGTTTGGT 66
|||:||||| ||| :|||:|||||:|||||: |||
Db 11 LysMetGluTyrArgLysSerGlyHisSerGlyLeuLysLeuProIleLeuSerLeuGly 30
|||:||||| ||| :|||:|||||:|||||: |||
QY 67 GCCTCTCGCTCGGAAGTGTCTTCGGTCCAGTCGCGGAAGATGATGCGTCGCCACCGCTG 126
|||:||||| ||| :|||:|||||:|||||: |||
Db 31 LeuTrpGlnAsnPheGlyAspTyrAspProIleHisAsnGlnArgGluIle-----Leu 48
|||:||||| ||| :|||:|||||:|||||: |||
QY 127 CGCAGGCTTTCCGTCCTCGGTATCAACTTCTTCGACACCTCCCGTATATTATGGAGGAACA 186
|||:||||| ||| :|||:|||||:|||||: |||
Db 49 ArgGlyAlaPheAspMetGlyIleThrHisPheAspLeuAlaAsnAsnTyrGlyGlyPro 68
|||:||||| ||| :|||:|||||:|||||: |||
QY 187 -----CTGCTCGAGAAATGCTTCGTAAAGGACTAAAGGCT---TTGCAAGTCCCTAGA 237
|||:||||| ||| :|||:|||||:|||||: |||
Db 69 AlaGlyAlaAlaGluLysAsnPheGlyArgIlePheArgGluAspPheGlnAlaTyrArg 88
|||:||||| ||| :|||:|||||:|||||: |||
QY 238 AGTGACTACATTCGTGCTTACTAAGTGTGCTAGATATAAAGAGGTTTTCATTTCAGTGCT 297
|||:||||| ||| :|||:|||||:|||||: |||
Db 89 AspGluLeuIleIleSerSerLysAlaGlyTyrHisMetTrpGluGlyProTyrGlyGlu 108
|||:||||| ||| :|||:|||||:|||||: |||
QY 298 GAGAGAGTAAGAAAGAGTATT-----GACGAGAGCTTCGAGAGGCTTCAGGCTT 345
|||:||||| ||| :|||:|||||:|||||: |||
Db 109 TrpGlySerArgLysSerIleIleSerSerCysAspGlnSerLeuGlnArgMetGlyLeu 128
|||:||||| ||| :|||:|||||:|||||: |||
QY 346 GATTATGTTGACATCTTCCATTCGCATGACATTGAGTTGCGGTCTCTTGATCAGATTGTG 405
|||:||||| ||| :|||:|||||:|||||: |||
Db 129 AspTyrValAspIlePheTyrHisHisArgProAsp-----ProAspThrProLeu 145
|||:||||| ||| :|||:|||||:|||||: |||
QY 406 AGTGAACAAATTCCTGCTCTTCAGAACTGGAACAGAGGGAACCCGGTTCATTGCT 465
|||:||||| ||| :|||:|||||:|||||: |||
Db 146 GluGluThrAlaGluAlaLeuMetGlnLeuValArgGlnGlyLysAlaLeuTyrIleGly 165
|||:||||| ||| :|||:|||||:|||||: |||
QY 466 ATCACT-----GGTCTTCCTCGGTAGATATTTTCACTATGTTCTTCGAGTGCCT 516
|||:||||| ||| :|||:|||||:|||||: |||
Db 166 IleSerAsnTyrAsnGluLysAspThrLysLysMetThrGluIleLeuLysArg----- 183
|||:||||| ||| :|||:|||||:|||||: |||
QY 517 CCAGGAGCTGCGATGATATTCATCTGCTCATTTAGCGGCTTAATGATTCGACGTTG 576
|||:||||| ||| :|||:|||||:|||||: |||
Db 184 ---LysGluAlaProPheIleIleHisGlnMetArgTyrAsnMetPheSerArgAlaLeu 202
|||:||||| ||| :|||:|||||:|||||: |||
QY 577 CTG-----GATTACTACTTACTTGAAGACAAAGGTGTGGGTGTCATAGTCTCTCT 630
|||:||||| ||| :|||:|||||:|||||: |||
Db 203 LeuGluAspAspLeuSerProValLeuGluGluGlyLeuGlyValAlaIleThrPheSer 222
|||:||||| ||| :|||:|||||:|||||: |||
QY 631 CCATTAGCAATGGGCTCTTACAGAACAA-----GGTCTCTCTGAATGGCACCT 681
|||:||||| ||| :|||:|||||:|||||: |||
Db 223 ProLeuAlaGlnGlyLeuLeuThrAsnArgTyrLeuHisGlyIleProGluAspSerArg 242
|||:||||| ||| :|||:|||||:|||||: |||
QY 682 GTTCC-----CCTGAG 693
|||:||||| ||| :|||:|||||:|||||: |||
Db 243 AlaHisArgLysGluIleProPheLeuSerGluGlnValGlySerThrLeuGluLys 262
|||:||||| ||| :|||:|||||:|||||: |||
QY 694 CTCAGTCTGCAAGCAAGCCAGTGTGCTCACTCAATCAAGGCAAGCAAGATCACA 753
|||:||||| ||| :|||:|||||:|||||: |||
Db 263 IleLysAlaLeuGlnThrIleAlaVal-----SerArgGlyGlnSerLeuAla 278
|||:||||| ||| :|||:|||||:|||||: |||
QY 754 AAGTTAGCTCTGCAATACAGTTTAGCAAAACAGAGAGATTTTCGTCGGTGTGTTGGGANG 813
|||:||||| ||| :|||:|||||:|||||: |||
Db 279 GlnMetAlaLeuAlaTrpAsnLeuArgGlnLysSerValThrSerValLeuValGlyAla 298
|||:||||| ||| :|||:|||||:|||||: |||
QY 814 ACCTCTGCTCAGGTAGAGAAATGTTGCGAGCAGTTACAGAGCTTGAAGTCTGGGG 873
|||:||||| ||| :|||:|||||:|||||: |||
Db 299 SerArgLeuSerGlnLeuGlnGluSerValArgMet-----MetAspAsnLeuAsp 315
|||:||||| ||| :|||:|||||:|||||: |||
QY 874 ATGGATCAAGAACTCTGCTCAGGTGTGAAGCTATTCTCGAG 915
|||:||||| ||| :|||:|||||:|||||: |||
Db 316 PheSerProGluGluLeuArgIleAspGlnIleLeuGlu 329
|||:||||| ||| :|||:|||||:|||||: |||
```

RESULT 15
ABO77872
ID ABO77872 standard; protein; 367 AA.
XX
AC ABO77872;

```
XX 29-JUL-2004 (first entry)  
XX Pseudomonas aeruginosa polypeptide #10047.  
XX Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.  
XX Pseudomonas aeruginosa.  
XX US6551795-B1.  
XX 22-APR-2003.  
XX 18-FEB-1999; 99US-00252991.  
XX 18-FEB-1998; 98US-0074788P.  
XX 27-JUL-1998; 98US-0094190P.  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX WPI; 2003-615309/58.  
XX N-PSDB; ABD11443.  
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide, of  
XX useful as molecular targets for diagnostics, prophylaxis and treatment of  
XX pathological conditions resulting from bacterial infection.  
XX Disclosure; SEQ ID NO 26618; 455pp; English.  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
XX polynucleotides encoding them. The sequences are useful in diagnosis and  
XX therapy of pathological conditions, as molecular targets for diagnostics,  
XX prophylaxis and treatment of pathological conditions resulting from a  
XX bacterial infection, for evaluating a compound, such as a polypeptide,  
XX for the ability to bind a P. aeruginosa nucleic acid, as components of  
XX effective antibacterial targets, as targets for antibacterial drugs,  
XX including anti-P. aeruginosa drugs, as templates for recombinant  
XX production of P. aeruginosa-derived peptides or polypeptides, as target  
XX components for diagnosis and/or treatment of P. aeruginosa-caused  
XX infection, and in detection of P. aeruginosa sequences or other sequences  
XX of Pseudomonas species using biochip technology. Sequences ABO67826-  
XX ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The  
XX sequence data for this patent did not form part of the printed  
XX specification but was obtained in electronic format from USPTO at  
XX seqdata.uspto.gov/sequence.html  
SQ Sequence 367 AA;
```

Alignment Scores:
Pred. No.: 1.12e-24 Length: 367
Score: 317.00 Matches: 84
Percent Similarity: 50.17% Conservative: 60
Best Local Similarity: 29.27% Mismatches: 103
Query Match: 19.18% Indels: 40
DB: Gaps: 8

US-10-606-300-12 (1-960) x ABO77872 (1-367)

```
QY 55 GTTGGTTTGGTCTCTCCGTCGGAAGTGTCTTCGGTCCAGTCGCCGAGATGATGCC 114  
|||:||||| ||| :|||:|||||:|||||: |||  
Db 52 IleGlyLeuGlyAlaProLeuGlyAsnMetPheHisProLeuSerGluGluThrAla 71  
|||:||||| ||| :|||:|||||:|||||: |||  
QY 115 GTCGCCACCGTCGCGAGGCTTTTCGTCCTCGGTATCAACTTCTTCGACACCTCCCGTAT 174  
|||:||||| ||| :|||:|||||:|||||: |||  
Db 72 AspAlaThrLeuAsnAlaAlaTrpAspAlaGlyPheArgTyrTyrAspValSerProHis 91  
|||:||||| ||| :|||:|||||:|||||: |||  
QY 175 TATGAGGAACACTGTCTGAGAAATGCTTGGTAGGGACTTAAGGCTTTGCAAGTCCCT 234  
|||:||||| ||| :|||:|||||:|||||: |||  
Db 92 TyrGlyAlaGlyLeuAlaGluGlnArgPheGlyArgLeuLeuSerGly-----LysPro 109  
|||:||||| ||| :|||:|||||:|||||: |||  
QY 235 AGAAGTGACTACATTGTGCTACTAAGTGTGCTAGATAT----- 273
```

	110	ArgAspGluTyrValLeuSerThrLysValGlyArgLeuLeuGlnProAlaSerGlnPro	129
	274	-----AAAGAAGGTTTGTATTCACGTGCT	297
	130	GluAsnAlaLysProPheValAspGluLeuProAsnLysArgValProAspTyrSerAla	149
	298	GAGAGAGTAAGAAAGATTGACGAGAGCTTGGAGAGCTTCACGTTGATTATGTTGAC	357
	150	AspGlyAlaArgArgSerIleGluAspSerLeuGluArgMetGlyValAspArgLeuAsp	169
	358	ATACTTCATTGCCATGACATT-----GAGTTCGGGTCT-----CTT	393
	170	ValValPheIleHisAspValSerGluAspGlnTrpGlyProGlnTrpArgGluTyrPhe	189
	394	GATCAGATTGTGAGTGAACAATTCCTGCTCTTCAGAACTCGAAACAGAGGGGAAGACC	453
	190	GlnGlnAlaMetAsnGlyAlaAlaLysAlaLeuThrGlnLeuArgAspGluGlyValIle	209
	454	CGGTTTCATTTGGTATCATCTGGTCTCCGGTTAGATATTTTCATCTATGTTCTTGATCGAGTG	513
	210	ArgGlyTrpGlyLeuGlyValAsnLeuValGluProCysArgLeuAlaLeuGluGlnSer	229
	514	CCTCCAGGAGCTGTCGATGTGATATTGTCATCTACTGTCATTACGGCGTT-----AATGAT	567
	230	AspPro-----AsnValPheLeuLeuAlaGlyArgTyrSerLeuLeuGluHisAsp	246
	568	TCGACGTTGCTGGATTACTACCTTACTTGAAGAGCAAGGTGGGTGTGATAGTGCT	627
	247	GluAlaLeuAspThrLeuPheProThrCysGlnAlaArgAspValGlyValValGly	266
	628	TCTCCATTAGCAATGGGCTCTTTCACAGAACAGGTCCTCTGTAATGGCACCCCTGCTTCC	687
	267	GlyProPheAsnSerGlyValLeuAlaGlyGlyAspHisTyrGluTyrAspGlnIlePro	286
	688	CCTGAGCTC-----AAGTCTCGAAGCAAGCCGAGTTGCTCACTGCAAAATCAAG	738
	287	ProGlnValAlaGlnArgArgGluGlnLeuLysAlaAlaGluHisCys-----	303
	739	GGCAAGAAGATCACAAAGTTAGCTCTGCAATACATGATTAGCAAAACAGAGGATTTTCGTCG	798
	304	GlyValAspLeuArgAlaAlaAlaLeuHisPheCysLeuAlaAsnProValValAlaSer	323
	799	GTGTTGTTGGGATGAGCTCT	819
	324	ValIleProGlyThrAlaAsn	330

Search completed: November 13, 2005, 08:26:25
Job time : 106 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2005, 04:30:28 ; Search time 622 Seconds
(without alignments)
9136.578 Million cell updates/sec

Title: US-10-606-300-12

Perfect score: 960

Sequence: 1 atgacgaaatagactctg.....gtggaatccatcagaactaa 960

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_16Dec04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001as:*
5: Geneseq2001bs:*
6: Geneseq2002as:*
7: Geneseq2002bs:*
8: Geneseq2003as:*
9: Geneseq2003bs:*
10: Geneseq2003cs:*
11: Geneseq2003ds:*
12: Geneseq2004as:*
13: Geneseq2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	960	100.0	960	4	AAD19526 Arabidops
2	960	100.0	960	6	Abk10125 A. thalia
3	946.4	98.6	1221	3	AAC37387 Arabidops
4	335	34.9	564	13	ACn48774 Cotton pr
5	224.8	23.4	563	13	ACn48848 Cotton pr
6	110	11.5	1038	4	ABL20521 Drosophil
7	98.2	10.2	1051	4	ABL28825 Drosophil
8	95.8	10.0	570	13	ACn47511 Cotton pr
9	89	9.3	484	13	ACn47619 Cotton pr
10	81.2	8.5	575	13	ACn46035 Cotton pr
11	80.2	8.4	885	4	ABL06603 Drosophil
12	71.2	7.4	3724	4	ABL20520 Drosophil
13	68.2	7.1	984	8	ACA38044 Prokaryot
14	56.6	5.9	3014	4	ABL06602 Drosophil
15	56.2	5.9	3308	4	ABL28824 Drosophil
16	56.2	5.9	10400	4	ABL20064 Drosophil
17	56.2	5.9	10418	4	ABL28814 Drosophil
18	56.2	5.9	10451	4	ABL19246 Drosophil
19	49.8	5.2	2000	8	ADA71938 Rice gene
20	46	4.8	586	13	ACn62765 Cotton ca

21	45	4.7	912	8	ACA21511	Aca21511 Prokaryot
22	42.4	4.4	960	13	ADT44646	Adt44646 Bacteria
23	42	4.4	975	13	ADS50132	Ads50132 Bacteria
24	41.6	4.3	912	6	ABN71096	Abn71096 Streptoco
25	41.6	4.3	918	6	ABN67863	Abn67863 Streptoco
26	41.6	4.3	110000	6	ABN71527_16	Continuation (17 o
27	40.4	4.2	1005	9	ADA30666	Ada30666 DNA encod
28	40	4.2	1104	11	ABD11443	Abd11443 Pseudomon
29	40	4.2	2514	11	ABD11187	Abd11187 Pseudomon
30	39.8	4.1	2679	2	AAQ28895	Aaq28895 Fucose de
31	39.4	4.1	927	13	ADS50023	Ads50023 Bacteria
32	39.4	4.1	927	13	ADS55539	Ads55539 Bacteria
33	39.2	4.1	987	2	AAQ62089	Aaq62089 L-fucose
34	39.2	4.1	1038	8	ACA31676	Aca31676 Prokaryot
35	39.2	4.1	2000	8	ADA71938	Ada71938 Rice gene
36	39	4.1	924	13	ADS55648	Ads55648 Bacteria
37	38.4	4.0	2000	12	ADJ41538	Adj41538 Plant cDN
38	38.2	4.0	650	13	ACn46597	Acn46597 Cotton pr
39	38	4.0	1452	8	ACA30734	Aca30734 Prokaryot
40	38	4.0	110000	2	RAX91990_04	Continuation (5 of
41	38	4.0	273254	3	AAC81914	Aac81914 Chlamydia
42	37.4	3.9	621	3	Aaf07906	Aaf07906 Fusarium
43	37.4	3.9	993	8	ACA43880	Aca43880 Prokaryot
44	37.4	3.9	999	8	ACA20791	Aca20791 Prokaryot
45	37.4	3.9	1512	4	AAF61103	Aaf61103 P. putida

ALIGNMENTS

RESULT 1

AAD19526	18-DEC-2001 (first entry)	Arabidopsis thaliana L-galactose dehydrogenase (L-galdH) cDNA.
ID	AAD19526 standard; cDNA; 960 BP.	L-galactose dehydrogenase; L-galdH; transgenic organism; ascorbic acid; herbicide; chromosome 4; ss.
XX		
AC	AAD19526;	
XX		
DT		
XX		
DB		
XX		
XX		
KW		
KX		
OS		
XX		
XX		
Key	Location/Qualifiers	
FT	CDS	1..960
FT		/*tag= a
FT		/product= "L-galactose dehydrogenase (L-galdH) "
XX		
XX		
PN		WO200172974-A2.
XX		
PD		04-OCT-2001.
XX		
XX		29-MAR-2001; 2001WO-GB001412.
XX		
XX		29-MAR-2000; 2000GB-00007651.
XX		(ASCO-) ASCORBEX LTD.
FA		
XX		
PI		Smirnoff N, Wheeler G;
XX		
DR		WPI; 2001-616482/71.
DR		P-PSDB; AAE11998.
XX		
PT		Novel L-galactose dehydrogenase protein and nucleic acid sequence encoding the protein for producing genetically modified plants and microorganisms with enhanced ability to synthesize ascorbic acid.
PT		
XX		
PS		Claim 6; Page 56-57; 58pp; English.
XX		
CC		The present invention relates to an isolated protein having L-galactose dehydrogenase (L-galdH) biological activity. L-galdH nucleic acid is
CC		

CC useful for generating transgenic organisms and modified pants with
CC enhanced ability to synthesize ascorbic acid. L-galDH facilitate the
CC production of a plant that has been genetically modified to express a
CC mutated L-galDH protein which is resistant to herbicides that act against
CC the naturally occurring L-galDH and to identify and/or design compounds
CC that are inhibitors of L-galDH. The compounds can be used, for e.g. in a
CC herbicide which acts on L-galDH and damages or kills plants that express
CC the enzyme. The present presence is Arabidopsis thaliana L-galactose
CC dehydrogenase (L-galDH) cDNA. The L-galDH gene is located on chromosome 4
XX
SQ Sequence 960 BP; 251 A; 195 C; 251 G; 263 T; 0 U; 0 Other;

Query Match 100.0%; Score 960; DB 4; Length 960;
Best Local Similarity 100.0%; Pred. No. 1.2e-304;
Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAAATAGAGCTTCGAGCTTTGGGGAACACAGGCTTAAGGTTAGCGCCGCTTGGT 60
DB 1 ATGACGAAATAGAGCTTCGAGCTTTGGGGAACACAGGCTTAAGGTTAGCGCCGCTTGGT 60
QY 61 TTTGGTGCTCTCCGCTCGAAGTGTCTTCGGTCCAGTCGCCGAAGATGATGCCGTCGCC 120
DB 61 TTTGGTGCTCTCCGCTCGAAGTGTCTTCGGTCCAGTCGCCGAAGATGATGCCGTCGCC 120
QY 121 ACCGTGCGGAGGCTTTCCGCTCCGATCACTTCTTCGACACCTCCCGTATTATGGA 180
DB 121 ACCGTGCGGAGGCTTTCCGCTCCGATCACTTCTTCGACACCTCCCGTATTATGGA 180
QY 181 GGAACACTGTCTGAGAAATGCTTCGTAAGGAGCTAAAGGCTTTGCAAGTCCCTAGAAGT 240
DB 181 GGAACACTGTCTGAGAAATGCTTCGTAAGGAGCTAAAGGCTTTGCAAGTCCCTAGAAGT 240
QY 241 GACTACATTTGGTGCTACTAAGTGTGGTATGATATAAAGAGGTTTTGATTTTCAGTCTCAG 300
DB 241 GACTACATTTGGTGCTACTAAGTGTGGTATGATATAAAGAGGTTTTGATTTTCAGTCTCAG 300
QY 301 AGAGTAAGAAAGATATTGACGAGAGCTTGGAGGCTTCAGCTTGATATTGTTGACATA 360
DB 301 AGAGTAAGAAAGATATTGACGAGAGCTTGGAGGCTTCAGCTTGATATTGTTGACATA 360
QY 361 CTTCAATTCGATGACATTTGAGTTCGGGTCTCTTGATCAGATTTGAGTGAACAATTCCT 420
DB 361 CTTCAATTCGATGACATTTGAGTTCGGGTCTCTTGATCAGATTTGAGTGAACAATTCCT 420
QY 421 GCTCTTCAGAACTGAAACAGAGGGGAAAGCCCGTTCAATGGTATCACTGGTCTTCGG 480
DB 421 GCTCTTCAGAACTGAAACAGAGGGGAAAGCCCGTTCAATGGTATCACTGGTCTTCGG 480
QY 481 TTAGATATTTTCACTATGTTCTTGATCGAGTGCCTCCAGGACCTGCGATGTGATATTG 540
DB 481 TTAGATATTTTCACTATGTTCTTGATCGAGTGCCTCCAGGACCTGCGATGTGATATTG 540
QY 541 TCATACCTGTCATTACGGGCTTAATGATTCGAGCTTGGCTGATTTACTACCTTACTTGAAG 600
DB 541 TCATACCTGTCATTACGGGCTTAATGATTCGAGCTTGGCTGATTTACTACCTTACTTGAAG 600
QY 601 AGCAAGGTGTGGGTGTGATAGTCTTCTCCATTAGCAATGGGCTCTCTACAGAACAA 660
DB 601 AGCAAGGTGTGGGTGTGATAGTCTTCTCCATTAGCAATGGGCTCTCTACAGAACAA 660
QY 661 GGTCTCTCTGAATGCGACCTCTCTCCCTGAGCTCAAGTCTGCAAGCAAGCCGAGTT 720
DB 661 GGTCTCTCTGAATGCGACCTCTCTCCCTGAGCTCAAGTCTGCAAGCAAGCCGAGTT 720
QY 721 GCTCTACTGCAATCAAGGGCAAGAGATCAAAAGTTAGCTCTGCAATACAGTTTAGCA 780
DB 721 GCTCTACTGCAATCAAGGGCAAGAGATCAAAAGTTAGCTCTGCAATACAGTTTAGCA 780
QY 781 AACAGGAGATTTCTGCGGTGTGGTGGGATGAGCTCTGCTCAAGTAGAAGAAAT 840
DB 781 AACAGGAGATTTCTGCGGTGTGGTGGGATGAGCTCTGCTCAAGTAGAAGAAAT 840
QY 841 GTTGACGAGCTTACAGAGCTTGAAAGTCTGGGGATGGAATCAAGAACTCTGCTGAGTT 900

DB 841 GTTGACGAGCTTACAGAGCTTGAAAGTCTGGGGATGGAATCAAGAACTCTGCTGAGTT 900
QY 901 GAAGCTATTTCGAGCCTGTAAAGAAATCTGACATGCGCAAGTGGGAATCCATCAGAACTAA 960
DB 901 GAAGCTATTTCGAGCCTGTAAAGAAATCTGACATGCGCAAGTGGGAATCCATCAGAACTAA 960
RESULT 2
ABK10125
ID ABK10125 standard; cDNA; 960 BP.
XX
AC ABK10125;
XX
DT 21-MAY-2002 (first entry)
XX
DE A. thaliana cDNA encoding L-galactose dehydrogenase (LGDH).
XX
KW ss; gene; ascorbic acid; vitamin C; scurvy; recombinant yeast;
KW L-galactose dehydrogenase; LGDH; L-galactono-1,4-lactone dehydrogenase;
KW AGD; D-arabinose dehydrogenase; ARA; D-arabinono-1,4-lactone oxidase;
KW ALO; L-gulono-1,4-lactone oxidase; GLO; aldonolactonase; AL.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 1..960
FT /*tag= a
FT /product= "LGDH"
XX
FT WO200210425-A2.
XX
PD 07-FEB-2002.
XX
PF 02-AUG-2001; 2001WO-GB003485.
XX
PR 02-AUG-2000; 2000US-00630983.
XX
PA (BIOP-) BIPOLO SCARL.
PA (WHAL/) WHALLEY K.
XX
PI Porro D, Sauer M;
XX
DR WPI; 2002-217125/27.
DR P-PSDB; AAU076343.
XX
PT Generating ascorbic acid or its salt, involves culturing yeast capable of
PT converting ascorbic acid precursor into ascorbic acid in medium
PT comprising ascorbic acid precursor, and isolating ascorbic acid.
XX
PS Claim 14; Page 87; 95pp; English.
XX
CC The invention relates to generating ascorbic acid or its salt, involves
CC obtaining a recombinant yeast capable of converting an ascorbic acid
CC precursor into ascorbic acid, culturing the recombinant yeast in a medium
CC comprising an ascorbic acid precursor, thus forming ascorbic acid, and
CC isolating the ascorbic acid. Also include are stabilising ascorbic acid
CC or its salt in a medium, by culturing a yeast in a medium comprising
CC ascorbic acid or its salt and a recombinant yeast functionally
CC transformed with a coding region encoding a protein having an enzyme
CC activity selected from L-galactose dehydrogenase (LGDH), L-galactono-1, 4
CC -lactone dehydrogenase (AGD), D-arabinose dehydrogenase (ARA), D-
CC arabinono-1,4-lactone oxidase (ALO), L-gulono-1,4-lactone oxidase (GLO)
CC and aldonolactonase (AL) activity, where the recombinant yeast is capable
CC of converting to ascorbic acid at least about 25% ascorbic acid precursor
CC or is capable of producing at least 20 mg ascorbic acid/L medium, when
CC the yeast is cultured in a medium comprising one ascorbic acid precursor.
CC The ascorbic acid produced (Vitamin C) is a powerful antioxidant, a
CC deficiency of which causes scurvy in humans. The present sequence encodes
CC A. thaliana LGDH
SQ Sequence 960 BP; 251 A; 195 C; 251 G; 263 T; 0 U; 0 Other;

Query Match		100.0%	Score 960	DB 6	Length 960	
Best Local Similarity		100.0%	Pred. No. 1.2e-304			
Matches 960		Conservative 0	Mismatches 0	Indels 0	Gaps 0	
QY	1	ATGACGAAATAGAGCTTCGAGCTTTGGGAAACACAGGCTTTAAGTTTAGCGCGTTCGT	60			
DB	1	ATGACGAAATAGAGCTTCGAGCTTTGGGAAACACAGGCTTTAAGTTTAGCGCGTTCGT	60			
QY	61	TTTGCTGCTCTCCGCTCGGAGTGTCTTCGGTCCAGTCCGCGAAGATGATGCCGTGCC	120			
DB	61	TTTGCTGCTCTCCGCTCGGAGTGTCTTCGGTCCAGTCCGCGAAGATGATGCCGTGCC	120			
QY	121	ACCGTCGCGAGGCTTTCGGTCTCGGTATCACTTCTCGACACCTCCCGTATTATGGA	180			
DB	121	ACCGTCGCGAGGCTTTCGGTCTCGGTATCACTTCTCGACACCTCCCGTATTATGGA	180			
QY	181	GGAACACTGTCTGAGAAATGCTTGGTAAGGACTTAAAGGCTTTGCAAGTCCCTAGAGT	240			
DB	181	GGAACACTGTCTGAGAAATGCTTGGTAAGGACTTAAAGGCTTTGCAAGTCCCTAGAGT	240			
QY	241	GACTACATTTGGCTTACTAAGTGTGGTAGATATATAAGAGGTTTTGATTTTCAGTGTGAG	300			
DB	241	GACTACATTTGGCTTACTAAGTGTGGTAGATATATAAGAGGTTTTGATTTTCAGTGTGAG	300			
QY	301	AGAGTAAGAAAGATATTGACGAGCTTTGAGAGGCTTTCAGTGTGATATTGTTGACATA	360			
DB	301	AGAGTAAGAAAGATATTGACGAGCTTTGAGAGGCTTTCAGTGTGATATTGTTGACATA	360			
QY	361	CTTCATTTCCCATGACATTTGAGTTCGGGCTCTTGATCAGATGTGAGTGAACAATTCCT	420			
DB	361	CTTCATTTCCCATGACATTTGAGTTCGGGCTCTTGATCAGATGTGAGTGAACAATTCCT	420			
QY	421	GCTCTTCAGAACTGAAACAGAGGGGAGACCCGGTTTCATTTGGTATCACTGGTCTTCGG	480			
DB	421	GCTCTTCAGAACTGAAACAGAGGGGAGACCCGGTTTCATTTGGTATCACTGGTCTTCGG	480			
QY	481	TTAGATATTTTCACTATTTGTTCTTGATCGAGTGCCTCCAGGACTGTGATATTG	540			
DB	481	TTAGATATTTTCACTATTTGTTCTTGATCGAGTGCCTCCAGGACTGTGATATTG	540			
QY	541	TCATCTGCTATAGGGGTTAATGATTCGAGTGTGCTGGATTTACTACCTTACTTTGAG	600			
DB	541	TCATCTGCTATAGGGGTTAATGATTCGAGTGTGCTGGATTTACTACCTTACTTTGAG	600			
QY	601	AGCAAGGTGGGTGTGATAGTCTCTCCATTAGCAATGGGCTCTTACAGAACAA	660			
DB	601	AGCAAGGTGGGTGTGATAGTCTCTCCATTAGCAATGGGCTCTTACAGAACAA	660			
QY	661	GGTCTCTGTAATGGCACCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAGCCGAGTT	720			
DB	661	GGTCTCTGTAATGGCACCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAGCCGAGTT	720			
QY	721	GCTCACTGCAATCAAGGGCAAGAGATCACAAGTTAGCTCTGCAATCAGTTTAGCA	780			
DB	721	GCTCACTGCAATCAAGGGCAAGAGATCACAAGTTAGCTCTGCAATCAGTTTAGCA	780			
QY	781	AACAGGAGATTTCTGCGGTGTGGTGGGATGAGCTCTGTCTCACAGGTAGAGAAAT	840			
DB	781	AACAGGAGATTTCTGCGGTGTGGTGGGATGAGCTCTGTCTCACAGGTAGAGAAAT	840			
QY	841	GTTGAGCAGTTTACAGAGCTTTGAAAGTCTGGGGATGATCAAGAACTCTGTCTGAGTT	900			
DB	841	GTTGAGCAGTTTACAGAGCTTTGAAAGTCTGGGGATGATCAAGAACTCTGTCTGAGTT	900			
QY	901	GAAGCTATTCGAGCCTGTAAGAACTGATGACATGGCCCAAGTGGATCCATCAGAACTAA	960			
DB	901	GAAGCTATTCGAGCCTGTAAGAACTGATGACATGGCCCAAGTGGATCCATCAGAACTAA	960			

RESULT 3
AAC37387
ID AAC37387 standard; DNA; 1221 BP.
XX

AC	AAC37387;	
XX		
DT	17-OCT-2000 (first entry)	
XX		
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 17196.	
XX		
KW	Hybridisation assay; genetic mapping; gene expression control;	
KW	protein identification; signal transduction pathway; metabolic pathway;	
KW	promoter; termination sequence; ss.	
XX		
OS	Arabidopsis thaliana.	
XX		
PN	EP1033405-A2.	
XX		
PD	06-SEP-2000.	
XX		
PF	25-FEB-2000; 2000EP-00301439.	
XX		
PR	25-FEB-1999; 99US-0121825P.	
PR	05-MAR-1999; 99US-0123180P.	
PR	09-MAR-1999; 99US-0123548P.	
PR	23-MAR-1999; 99US-0125788P.	
PR	25-MAR-1999; 99US-0126264P.	
PR	29-MAR-1999; 99US-0126785P.	
PR	01-APR-1999; 99US-0127462P.	
PR	06-APR-1999; 99US-0128234P.	
PR	08-APR-1999; 99US-0128714P.	
PR	16-APR-1999; 99US-0129845P.	
PR	19-APR-1999; 99US-0130077P.	
PR	21-APR-1999; 99US-0130449P.	
PR	23-APR-1999; 99US-0130510P.	
PR	23-APR-1999; 99US-0130891P.	
PR	30-APR-1999; 99US-0131449P.	
PR	30-APR-1999; 99US-0132048P.	
PR	30-APR-1999; 99US-0132407P.	
PR	04-MAY-1999; 99US-0132484P.	
PR	05-MAY-1999; 99US-0132485P.	
PR	06-MAY-1999; 99US-0132486P.	
PR	06-MAY-1999; 99US-0132487P.	
PR	07-MAY-1999; 99US-0132863P.	
PR	11-MAY-1999; 99US-0134256P.	
PR	14-MAY-1999; 99US-0134218P.	
PR	14-MAY-1999; 99US-0134219P.	
PR	14-MAY-1999; 99US-0134221P.	
PR	14-MAY-1999; 99US-0134370P.	
PR	18-MAY-1999; 99US-0134768P.	
PR	19-MAY-1999; 99US-0134941P.	
PR	20-MAY-1999; 99US-0135124P.	
PR	21-MAY-1999; 99US-0135353P.	
PR	24-MAY-1999; 99US-0135629P.	
PR	25-MAY-1999; 99US-0136021P.	
PR	27-MAY-1999; 99US-0136392P.	
PR	28-MAY-1999; 99US-0136782P.	
PR	01-JUN-1999; 99US-0137222P.	
PR	03-JUN-1999; 99US-0137528P.	
PR	04-JUN-1999; 99US-0137502P.	
PR	07-JUN-1999; 99US-0137724P.	
PR	08-JUN-1999; 99US-0138094P.	
PR	10-JUN-1999; 99US-0138540P.	
PR	10-JUN-1999; 99US-0138847P.	
PR	14-JUN-1999; 99US-0139119P.	
PR	16-JUN-1999; 99US-0139452P.	
PR	16-JUN-1999; 99US-0139453P.	
PR	17-JUN-1999; 99US-0139492P.	
PR	18-JUN-1999; 99US-0139454P.	
PR	18-JUN-1999; 99US-0139455P.	
PR	18-JUN-1999; 99US-0139456P.	
PR	18-JUN-1999; 99US-0139457P.	
PR	18-JUN-1999; 99US-0139458P.	
PR	18-JUN-1999; 99US-0139459P.	
PR	18-JUN-1999; 99US-0139460P.	
PR	18-JUN-1999; 99US-0139461P.	
PR	18-JUN-1999; 99US-0139462P.	

PR 18-JUN-1999;	99US-0139463P.	PR 26-AUG-1999;	99US-0150884P.
PR 18-JUN-1999;	99US-0139750P.	PR 27-AUG-1999;	99US-0151065P.
PR 18-JUN-1999;	99US-0139763P.	PR 27-AUG-1999;	99US-0151066P.
PR 21-JUN-1999;	99US-0139817P.	PR 27-AUG-1999;	99US-0151080P.
PR 22-JUN-1999;	99US-0139899P.	PR 30-AUG-1999;	99US-0151303P.
PR 23-JUN-1999;	99US-0140353P.	PR 31-AUG-1999;	99US-0151438P.
PR 23-JUN-1999;	99US-0140354P.	PR 01-SEP-1999;	99US-0151930P.
PR 24-JUN-1999;	99US-0140695P.	PR 07-SEP-1999;	99US-0152363P.
PR 28-JUN-1999;	99US-0140823P.	PR 10-SEP-1999;	99US-0153070P.
PR 29-JUN-1999;	99US-0140991P.	PR 13-SEP-1999;	99US-0153758P.
PR 30-JUN-1999;	99US-0141287P.	PR 15-SEP-1999;	99US-0154018P.
PR 01-JUL-1999;	99US-0141843P.	PR 16-SEP-1999;	99US-0154039P.
PR 01-JUL-1999;	99US-0142154P.	PR 20-SEP-1999;	99US-0154779P.
PR 02-JUL-1999;	99US-0142055P.	PR 22-SEP-1999;	99US-0155139P.
PR 06-JUL-1999;	99US-0142390P.	PR 23-SEP-1999;	99US-0155486P.
PR 08-JUL-1999;	99US-0142803P.	PR 24-SEP-1999;	99US-0155659P.
PR 09-JUL-1999;	99US-0142920P.	PR 28-SEP-1999;	99US-0156458P.
PR 12-JUL-1999;	99US-0142977P.	PR 29-SEP-1999;	99US-0156596P.
PR 13-JUL-1999;	99US-0143542P.	PR 04-OCT-1999;	99US-0157117P.
PR 14-JUL-1999;	99US-0143624P.	PR 05-OCT-1999;	99US-0157753P.
PR 15-JUL-1999;	99US-0144005P.	PR 06-OCT-1999;	99US-0157865P.
PR 16-JUL-1999;	99US-0144085P.	PR 07-OCT-1999;	99US-0158029P.
PR 16-JUL-1999;	99US-0144086P.	PR 08-OCT-1999;	99US-0158232P.
PR 19-JUL-1999;	99US-0144325P.	PR 12-OCT-1999;	99US-0158369P.
PR 19-JUL-1999;	99US-0144331P.	PR 13-OCT-1999;	99US-0159293P.
PR 19-JUL-1999;	99US-0144332P.	PR 13-OCT-1999;	99US-0159294P.
PR 19-JUL-1999;	99US-0144333P.	PR 13-OCT-1999;	99US-0159295P.
PR 19-JUL-1999;	99US-0144334P.	PR 14-OCT-1999;	99US-0159329P.
PR 19-JUL-1999;	99US-0144335P.	PR 14-OCT-1999;	99US-0159331P.
PR 20-JUL-1999;	99US-0144352P.	PR 14-OCT-1999;	99US-0159637P.
PR 20-JUL-1999;	99US-0144632P.	PR 14-OCT-1999;	99US-0159638P.
PR 21-JUL-1999;	99US-0144884P.	PR 18-OCT-1999;	99US-0159584P.
PR 21-JUL-1999;	99US-0144814P.	PR 21-OCT-1999;	99US-0160741P.
PR 21-JUL-1999;	99US-0145086P.	PR 21-OCT-1999;	99US-0160767P.
PR 22-JUL-1999;	99US-0145085P.	PR 21-OCT-1999;	99US-0160768P.
PR 22-JUL-1999;	99US-0145087P.	PR 21-OCT-1999;	99US-0160770P.
PR 22-JUL-1999;	99US-0145089P.	PR 21-OCT-1999;	99US-0160814P.
PR 22-JUL-1999;	99US-0145192P.	PR 21-OCT-1999;	99US-0160815P.
PR 23-JUL-1999;	99US-0145145P.	PR 22-OCT-1999;	99US-0160980P.
PR 23-JUL-1999;	99US-0145218P.	PR 22-OCT-1999;	99US-0160981P.
PR 23-JUL-1999;	99US-0145224P.	PR 22-OCT-1999;	99US-0160989P.
PR 26-JUL-1999;	99US-0145276P.	PR 25-OCT-1999;	99US-0161404P.
PR 27-JUL-1999;	99US-0145913P.	PR 25-OCT-1999;	99US-0161405P.
PR 27-JUL-1999;	99US-0145918P.	PR 25-OCT-1999;	99US-0161406P.
PR 27-JUL-1999;	99US-0145919P.	PR 26-OCT-1999;	99US-0161359P.
PR 28-JUL-1999;	99US-0145951P.	PR 26-OCT-1999;	99US-0161360P.
PR 02-AUG-1999;	99US-0146386P.	PR 26-OCT-1999;	99US-0161361P.
PR 02-AUG-1999;	99US-0146388P.	PR 28-OCT-1999;	99US-0161920P.
PR 02-AUG-1999;	99US-0146389P.	PR 28-OCT-1999;	99US-0161992P.
PR 04-AUG-1999;	99US-0147038P.	PR 28-OCT-1999;	99US-0161993P.
PR 04-AUG-1999;	99US-0147204P.	PR 29-OCT-1999;	99US-0162142P.
PR 05-AUG-1999;	99US-0147192P.		
PR 05-AUG-1999;	99US-0147260P.		
PR 06-AUG-1999;	99US-0147303P.		
PR 06-AUG-1999;	99US-0147416P.		
PR 09-AUG-1999;	99US-0147493P.		
PR 09-AUG-1999;	99US-0147935P.		
PR 10-AUG-1999;	99US-0148171P.		
PR 11-AUG-1999;	99US-0148319P.		
PR 12-AUG-1999;	99US-0148341P.		
PR 13-AUG-1999;	99US-0148565P.		
PR 13-AUG-1999;	99US-0148684P.		
PR 16-AUG-1999;	99US-0149368P.		
PR 17-AUG-1999;	99US-0149175P.		
PR 18-AUG-1999;	99US-0149426P.		
PR 20-AUG-1999;	99US-0149722P.		
PR 20-AUG-1999;	99US-0149723P.		
PR 20-AUG-1999;	99US-0149929P.		
PR 23-AUG-1999;	99US-0149902P.		
PR 23-AUG-1999;	99US-0149930P.		
PR 25-AUG-1999;	99US-0150566P.		

Query Match

Best Local Similarity 98.6%; Score 946.4; DB 3; Length 1221;

Matches 958; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY	1	ATGACGAAAAATAGAGCTTCGAGCTTTGGGGAACACAGGGCTTTAAGGTTAGCGCCGTTGGT	60
Db	116	ATGACGAAAAATAGAGCTTC-AGCTTTGGGGAACACAGGGCTTTAAGGTTAGCGCCGTTGGT	174
QY	61	TTTGGTGCCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCCGGAAGATGATCGCGTCGCC	120
Db	175	TTTGGTGCCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCCGGAAGATGATCGCGTCGCC	234
QY	121	ACCGTGCAGGAGCTTTCCGCTCTCGGTATCAACTTCTTCGACACCTCCCGTATTATGGA	180
Db	235	ACCGTGCAGGAGCTTTCCGCTCTCGGTATCAACTTCTTCGACACCTCCCGTATTATGGA	294
QY	181	GGACACACTGTCTGAGAAAAATGCTTGGTAAGGGA-CTAAAGGCTTTGCAAGTCCCTAGAACT	240
Db	295	GGACACACTGTCTGAGAAAAATGCTTGGTAAGGGA-CTAAAGGCTTTGCAAGTCCCTAGAACT	354

QY 241 GACTACATTGGCTACTAAGTGTGGTAGATATAAAGAGGTTTGTGATTTTCAGTCTGAG 300
DB |||||||
QY 355 GACTACATTGGCTACTAAGTGTGGTAGATATAAAGAGGTTTGTGATTTTCAGTCTGAG 414
DB |||||||
QY 301 AGAGTAAGAAAGAGTATTGACGAGAGCTTGGAGAGGCTTCAGCTTGATTTGTCACATA 360
DB |||||||
QY 415 AGAGTAAGAAAGAGTATTGACGAGAGCTTGGAGAGGCTTCAGCTTGATTTGTCACATA 474
DB |||||||
QY 361 TTTCATTGCCATGACATTGAGTTGGGTTCTCTTGATCAGATTGTGAGTGAACAATTCCT 420
DB |||||||
QY 475 TTTCATTGCCATGACATTGAGTTGGGTTCTCTTGATCAGATTGTGAGTGAACAATTCCT 534
DB |||||||
QY 421 GCTCTTCAGAACTGAACAGAGGGGAACCCGGTTCATTTGATCAGTGTCTTCG 480
DB |||||||
QY 535 GCTCTTCAGAACTGAACAGAGGGGAACCCGGTTCATTTGATCAGTGTCTTCG 594
DB |||||||
QY 481 TTAGATATTTTTCACATTATGTTCTTCATCGAGTGCCTCCAGGAGCTTCGATGTGATTTG 540
DB |||||||
QY 595 TTAGATATTTTTCACATTATGTTCTTCATCGAGTGCCTCCAGGAGCTTCGATGTGATTTG 654
DB |||||||
QY 541 TCATACCTGTCAATTACGGGTTAATGATTCGAGCTTGTGGATTTACTACTTACTTTGAAG 600
DB |||||||
QY 655 TCATACCTGTCAATTACGGGTTAATGATTCGAGCTTGTGGATTTACTACTTACTTTGAAG 714
DB |||||||
QY 601 AGCAAGGTGGGTGTGATAGTCTTCTTCATAGCAATGGGCTCTCTTACAGAACAA 660
DB |||||||
QY 715 AGCAAGGTGGGTGTGATAGTCTTCTTCATAGCAATGGGCTCTCTTACAGAACAA 774
DB |||||||
QY 661 GGTCTCTCGAATGACCTGCTTCCCTCGAGTCAAGTCTGCAAGCAAGGCGCAGTT 720
DB |||||||
QY 775 GGTCTCTCGAATGACCTGCTTCCCTCGAGTCAAGTCTGCAAGCAAGGCGCAGTT 834
DB |||||||
QY 721 GCTCACTGCAATCAAGGCGCAAGAGATCAAAAGTTAGCTTCTGCAATACAGTTTAGCA 780
DB |||||||
QY 835 GCTCACTGCAATCAAGGCGCAAGAGATCAAAAGTTAGCTTCTGCAATACAGTTTAGCA 894
DB |||||||
QY 781 AACAAAGAGATTTCTCGTGTGTTGGATGAGCTCTGTCTCACAGTGAAGAAAT 840
DB |||||||
QY 895 AACAAAGAGATTTCTCGTGTGTTGGATGAGCTCTGTCTCACAGTGAAGAAAT 954
DB |||||||
QY 841 GTTGACAGAGTTACAGAGCTTGAAGCTTGGGATGGATCAAGAACTCTGCTGAGTT 900
DB |||||||
QY 955 GTTGACAGAGTTACAGAGCTTGAAGCTTGGGATGGATCAAGAACTCTGCTGAGTT 1014
DB |||||||
QY 901 GAAGCTATTCTGAGCTGTAAAGAACTGCAATGCGCAAGTGAATCCATCAGAACTAA 960
DB |||||||
QY 1015 GAAGCTATTCTGAGCTGTAAAGAACTGCAATGCGCAAGTGAATCCATCAGAACTAA 1074
DB |||||||

RESULT 4

ACN48774
ID ACN48774 standard; cDNA; 564 BP.
XX AC ACN48774;
XX DT 02-DEC-2004 (first entry)
XX DE Cotton primed seed EST Clone ID: LIB3825-027-Q6-K6-G2, SEQ:3555.
XX KW Cotton; plant; EST; expressed sequence tag; transgenic plant; seed;
XX KW variety DP50B; library LIB3825; molecular tag; molecular marker;
XX KW genetic mapping; molecular mapping; seed germination; plant growth;
XX KW plant quality; plant yield; plant breeding; tissue printing; ss.
OS Gossypium hirsutum.
XX US2004123340-A1.
XX PN
XX PD 24-JUN-2004.
XX PF 12-DEC-2001; 2001US-00021323.
XX PP 14-DEC-2000; 2000US-0255619P.

XX (DEIK/) DEIKMAN J.
PA (FENG/) FENG P C C.
PA (FENG/) FENG P C C.
PA (FENG/) FENG P C C.
XX (ZIEG/) ZIEGLER T E.
PI Deikman J, Feng PCC, Fincher KL, Ziegler TE;
XX WPI; 2004-479808/45.

New isolated nucleic acid molecule that encodes a plant protein or its fragment, useful for isolating a variety of agronomically significant genes associated with plant growth, quality or yield, and as molecular tags to map genes.

Claim 1; SEQ ID NO 3555; 34pp; English.

The invention relates to 17880 cotton expressed sequence tags (ESTs; ACN45220-ACN63099). The ESTs were isolated from CDNA libraries generated from primed or non-primed seeds from variety DP50B, mature seeds from variety Coker 312 Boswell 96 Field, and androecium tissue, gynoecium tissue, developing fibres, carpel walls and septa from variety Nucotton33B. The invention also relates to substantially purified proteins or their fragments encoded by nucleic acid molecules of the invention, and to transformed plants having a nucleic acid construct comprising a nucleic acid of the invention. The cotton ESTs are useful as molecular tags to isolate genetic regions, to isolate genes, to map genes, to determine gene function and to determining whether genes are members of a particular gene family. The nucleic acid molecules may be used for isolating a variety of agronomically significant genes associated with plant growth, quality, yield, and could also serve as links in metabolic and catabolic pathways. The nucleic acid molecules are also useful for identifying genes important in initiating and maintaining seed germination or that may be used to mitigate stresses encountered during seed germination. The ESTs additionally enable the acquisition of promoters and cis-regulatory elements which will be useful to express agronomically significant genes in these tissues and/or other tissues, and also permits the acquisition of molecular markers useful in breeding schemes, genetic and molecular mapping, and in cloning of agronomically significant genes. The nucleic acid molecules are further useful for detecting the expression level or pattern of a protein or mRNA and for detecting the presence or quantity of a protein by tissue printing. The present sequence represents a specifically claimed EST isolated from a cotton variety DP50B primed seed CDNA library (LIB3825). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the US patent office at seqdata.uspto.gov/sequence.html?DocID=US20040123340

XX SQ Sequence 564 BP; 156 A; 97 C; 135 G; 175 T; 0 U; 1 Other;

Query Match 34.9%; Score 335; DB 13; Length 564;
Best Local Similarity 77.0%; Pred. No. 5e-99;
Matches 434; Conservative 0; Mismatches 126; Indels 4; Gaps 2;

QY 198 AATGCTTGGTAAGGACCTAAAGGCTTTCAGAGTCCCTAGAGTGAAGTGTGCTGCTAC 257
DB |||||||
QY 258 TAAGTGTGTAGATATAAAGAGGTTTGTGATTTTCAGTCTGAGAGGTAAGAGAGTAT 317
DB |||||||
QY 62 CAATGTGGGAGATATCTGTGAAGGTTTGTGATTTTCAGTCTGAGAGGTAAGAGTAT 121
DB |||||||
QY 318 TGACGAGAGCTTGGAGAGGCTTTCAGTCTGATTTGATTTGACATACATTTTCATGCAAT 377
DB |||||||
QY 122 TGATGAAGCTTGGAGAGGTTTTCAGTCTGATTTGATTTGATTTTCATGCAATGAT 181
DB |||||||
QY 378 TGAGTTCGGGCTCTCTTGATCAGATTGTGAGTGAACAATTCCTGCTCTTCAGAACTGAA 437
DB |||||||
QY 182 TGAATTCGGCTCTCTTGATCAGATTGTGAGTGAACAATTCCTGCTCTTCAGAACTGAA 241
DB |||||||
QY 438 ACAAGAGGGGAAGACCCGGTTCATTTGATCAGTGTCTTCGTTAGATATTTTCACTTA 497
DB |||||||
QY 242 GGAAGCAGGGAAGATTCTGTTTTCATTTGGTATCACCGGGTTCCTTGGAAATTTTACTTA 301
DB |||||||


```
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 13036; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1038 BP; 263 A; 257 C; 304 G; 214 T; 0 U; 0 Other;
XX
Query Match 11.5%; Score 110; DB 4; Length 1038;
Best Local Similarity 50.9%; Pred. No. 8.5e-25;
Matches 344; Conservative 0; Mismatches 320; Indels 12; Gaps 3;
QY 106 GATGATCGCGCCGACCGTGGCGAGGCTTCGCTCGGTATCAACTTCTTCGACACC 165
DB 157 GAGGAGGGTATTAACCTGTGCAGGCGCGTAAAGTCAGGCATCAACTACATTGACACT 216
QY 166 TCCCGTATTATGGAGGACACTGCTGAGAAATGCTTGTGAAGGACTAAGCCTTGG 225
DB 217 GCTCCCTGGTATGTCAGGGTCGCTGAGGAGGTCCTGGGACTGCGCCCTAAGGATGTG 276
QY 226 C-----AAGTCCCTAGAGTACATCTTGGGCTACTAAGTGTGTAGATATAAGAA 279
DB 277 CCGCGGGATCTACTATATGCCACGAAAGTGGCTCGTACGAACTGACTACGATAA 336
QY 280 GGTTTTGATTTGAGTGGAGAGATAGAAAGATATTTGACGAGCTTGGAGGGTTT 339
DB 337 ATGTTTGACTTTAGTGCCCAAGAGACGCGGAAAGCGTGGAGAGCTTGAACACTCTT 396
QY 340 CAGCTTGATTTGACATCTTCAATGCCATGACATGATTTGG---GGTCTTTGAT 396
DB 397 GCGCTGGACTAGTTGATGTCTATCCAGATTCACGATTCGAGTTCCGCAAGATCTGGAC 456
QY 397 CAGATTGTGAGTGAACAATCTCTCTTCAGAACTGMAACAAGAGGGAAGACCGG 456
DB 457 ATTGTGATCAAGAGACACTGCCCACTTTGGAGAGCTGTGTCAGAGGGGCAAGCAAGA 516
QY 457 TTCAATGGTATCAGTGGTCTTCGTTAGATATTTTCACTTATGTTTTCATGAGTGGCT 516
DB 517 TTCAATGGAGTGTCCGCTTACCGATTTCCGGTCTTAAGGAGTTCTTGACCGGA---ACA 573
QY 517 CAGGGAGCTGTCGATGTGATATGTCATCTGATAGGGGTTTAATGATTCGAGCTTG 576
DB 574 GCGGGAAGACTCGATACGCTCTCACCTATGTCAGATACACCTGACCGATGAACGCTC 633
QY 577 CTGGATTACTACTTACTTGAAGCAAGAGTGTGGGTGTGATAGTGTCTTCCATT 636
DB 634 CTGGAGTACCTGGATTTCTTCAAGTCCCAAGAACCTTGGGCGTCACTGTGCGCGAGCTCAT 693
QY 637 GCAATGGGCTCTTACAGAACAAAGTCTCTCTGAATGGCAACCTCTGCTTCCCTGAGCTC 696
```

```

694 GCCCTCGACTGCTGACCAATGCGCGTCCACAGGCATGGCATCCGCGCAGTGATGAGCAG 753
697 AAGTCTGCAAGCAAAAGCGCAGATTGCTCCTCCTCAATCAAAAGGCGCAAGAGATCAAAAG 756
754 AAGGCCATTGCGCCGGAAGGCATCGGAGGCTCTGCAAGGAACGCGGCTGGAGCTGGGCAAG 813
757 TTAGCTCTGCAATACA 772
814 CTGGCCATGTACTACA 829

RESULT 7
ABL28825
ID ABL28825 standard; DNA; 1051 BP.
XX
XX ABL28825;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 37948.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 37948; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1051 BP; 242 A; 271 C; 305 G; 233 T; 0 U; 0 Other;
XX
Query Match 10.2%; Score 98.2; DB 4; Length 1051;
Best Local Similarity 49.9%; Pred. No. 6.6e-21;
Matches 359; Conservative 0; Mismatches 348; Indels 12; Gaps 4;
QY 63 TGGTGCTCTCCGCTCGGAAGTGTCTTCGCTCCAGTCCCGAAGATGATGCGTGCAC 122
DB 142 TGGTGCCACTCTCTCCAGCTCTTCTCCGATGACTTCTCCGAGGAGGCGATCCTCAC 201
QY 123 CGTGGCGAGCTTCCGCTCTCGGTATCACTTCTCGACACCTCCCGTATTATGGAGG 182
DB 202 GGTGCGAAGGCCATTAGATCCGGTATCACTACATACAGACGGCTCCCTTTATGGCCA 261
```


XX Gossypium hirsutum.
XX US2004123340-A1.
XX 24-JUN-2004.
XX 12-DEC-2001; 2001US-00021323.
XX 14-DEC-2000; 2000US-0255619P.
XX (DEIK/) DEIKMAN J.
XX (FENG/) FENG P C C.
XX (FINC/) FINCHER K L.
XX (ZIEG/) ZIEGLER T E.
XX Deikman J, Feng PCC, Fincher KL, Ziegler TE;
XX WPI; 2004-479808/45.
XX New isolated nucleic acid molecule that encodes a plant protein or its
XX fragment, useful for isolating a variety of agronomically significant
XX genes associated with plant growth, quality or yield, and as molecular
XX tags to map genes.
XX Claim 1; SEQ ID NO 2400; 34pp; English.
XX The invention relates to 17880 cotton expressed sequence tags (ESTs;
XX ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated
XX from primed or non-primed seeds from variety DF50B, mature seeds from
XX variety Coker 312 Boswell 96 Field, and androecium tissue, gynoecium
XX tissue, developing fibres, carpel walls and septa from variety
XX Nucotton33B. The invention also relates to substantially purified
XX proteins or their fragments encoded by nucleic acid molecules of the
XX invention, and to transformed plants having a nucleic acid construct
XX comprising a nucleic acid of the invention. The cotton ESTs are useful as
XX molecular tags to isolate genetic regions, to isolate genes, to map
XX genes, to determine gene function and to determining whether genes are
XX members of a particular gene family. The nucleic acid molecules may be
XX used for isolating a variety of agronomically significant genes
XX associated with plant growth, quality, yield, and could also serve as
XX links in metabolic and catabolic pathways. The nucleic acid molecules are
XX also useful for identifying genes important in initiating and maintaining
XX seed germination or that may be used to mitigate stresses encountered
XX during seed germination. The ESTs additionally enable the acquisition of
XX agronomically significant genes in these tissues and/or other tissues,
XX and also permits the acquisition of molecular markers useful in breeding
XX schemes, genetic and molecular mapping, and in cloning of agronomically
XX significant genes. The nucleic acid molecules are further useful for
XX detecting the expression level or pattern of a protein or mRNA and for
XX present sequence represents a specifically claimed EST isolated from a
XX cotton variety DF50B primed seed cDNA library (LIB3825). The sequence
XX data for this patent did not form part of the printed specification, but
XX was obtained in electronic format directly from the US patent office at
XX seqdata.uspto.gov/sequence.html?DocID=US20040123340
XX
XX Sequence 484 BP; 128 A; 121 C; 93 G; 142 T; 0 U; 0 Other;
XX
XX Query Match 9.3%; Score 89; DB 13; Length 484;
XX Best Local Similarity 73.9%; Pred. No. 4.4e-18;
XX Matches 113; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
XX
XX 20 GAGCTTGGGACACAGGCTTAGGTTAGCGCGTGGTGGTGGCTTCCGCTCG 79
XX 235 GACCTCTCGGAACACGCGCCCTCAAGCTCAGCAGCGTGGGCTTCTCTCTCG 294
XX 80 GAAGTGTCTTCGGTCCAGTCCGCGGAGATGATGCCCTGCCACCGCGCGAGGCTTCC 139
XX 295 GCAGCGCTCTTAGTTCGGTGGCGAAGCAGCGTGTGCGCTCCGCTCGGAGGCTATC 354
XX 140 GTCTCGGTATCAACTTCTTCGACACCTCCCGT 172

Db 355 GCGTCGAATCAACTTCTTCGACACCTCTACGT 387

RESULT 10

ACN46035/c

ID ACN46035 standard; cDNA; 575 BP.

XX ACN46035;

XX 02-DEC-2004 (first entry)

XX Cotton primed seed EST Clone ID: LIB3825-013-Q1-N6-E6, SEQ:816.

XX Cotton; plant; EST; expressed sequence tag; transgenic plant; seed;

XX variety DF50B; library LIB3825; molecular tag; molecular marker;

XX genetic mapping; molecular mapping; seed germination; plant growth;

XX plant quality; plant yield; plant breeding; tissue printing; ss.

XX Gossypium hirsutum.

XX US2004123340-A1.

XX 24-JUN-2004.

XX 12-DEC-2001; 2001US-00021323.

XX 14-DEC-2000; 2000US-0255619P.

XX (DEIK/) DEIKMAN J.

XX (FENG/) FENG P C C.

XX (FINC/) FINCHER K L.

XX (ZIEG/) ZIEGLER T E.

XX Deikman J, Feng PCC, Fincher KL, Ziegler TE;

XX WPI; 2004-479808/45.

XX New isolated nucleic acid molecule that encodes a plant protein or its

XX fragment, useful for isolating a variety of agronomically significant

XX genes associated with plant growth, quality or yield, and as molecular

XX tags to map genes.

XX Claim 1; SEQ ID NO 816; 34pp; English.

XX The invention relates to 17880 cotton expressed sequence tags (ESTs;
XX ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated
XX from primed or non-primed seeds from variety DF50B, mature seeds from
XX variety Coker 312 Boswell 96 Field, and androecium tissue, gynoecium
XX tissue, developing fibres, carpel walls and septa from variety
XX Nucotton33B. The invention also relates to substantially purified
XX proteins or their fragments encoded by nucleic acid molecules of the
XX invention, and to transformed plants having a nucleic acid construct
XX comprising a nucleic acid of the invention. The cotton ESTs are useful as
XX molecular tags to isolate genetic regions, to isolate genes, to map
XX genes, to determine gene function and to determining whether genes are
XX members of a particular gene family. The nucleic acid molecules may be
XX used for isolating a variety of agronomically significant genes
XX associated with plant growth, quality, yield, and could also serve as
XX links in metabolic and catabolic pathways. The nucleic acid molecules are
XX also useful for identifying genes important in initiating and maintaining
XX seed germination or that may be used to mitigate stresses encountered
XX during seed germination. The ESTs additionally enable the acquisition of
XX promoters and cis-regulatory elements which will be useful to express
XX agronomically significant genes in these tissues and/or other tissues,
XX and also permits the acquisition of molecular markers useful in breeding
XX schemes, genetic and molecular mapping, and in cloning of agronomically
XX significant genes. The nucleic acid molecules are further useful for
XX detecting the expression level or pattern of a protein or mRNA and for
XX present sequence represents a specifically claimed EST isolated from a
XX cotton variety DF50B primed seed cDNA library (LIB3825). The sequence
XX data for this patent did not form part of the printed specification, but
XX was obtained in electronic format directly from the US patent office at
XX seqdata.uspto.gov/sequence.html?DocID=US20040123340
XX

XX Sequence 484 BP; 128 A; 121 C; 93 G; 142 T; 0 U; 0 Other;

XX Query Match 9.3%; Score 89; DB 13; Length 484;
XX Best Local Similarity 73.9%; Pred. No. 4.4e-18;
XX Matches 113; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
XX
XX 20 GAGCTTGGGACACAGGCTTAGGTTAGCGCGTGGTGGTGGCTTCCGCTCG 79
XX 235 GACCTCTCGGAACACGCGCCCTCAAGCTCAGCAGCGTGGGCTTCTCTCTCG 294
XX 80 GAAGTGTCTTCGGTCCAGTCCGCGGAGATGATGCCCTGCCACCGCGCGAGGCTTCC 139
XX 295 GCAGCGCTCTTAGTTCGGTGGCGAAGCAGCGTGTGCGCTCCGCTCGGAGGCTATC 354
XX 140 GTCTCGGTATCAACTTCTTCGACACCTCCCGT 172

CC was obtained in electronic format directly from the US patent office at
CC seqdata.uspto.gov/sequence.html?DocID=US20040123340
XX
SQ Sequence 575 BP; 187 A; 121 C; 99 G; 168 T; 0 U; 0 Other;
Query Match 8.5%; Score 81.2; DB 13; Length 575;
Best Local Similarity 75.4%; Pred. No. 1.8e-15;
Matches 101; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 827 AGGTAGAGAAAATGTTGCAGCAGTTACAGAGCTTTGAAAGTCTGGGGATGGATCAAGAAA 886
DB 417 AGGTTGAAGAGAAATGTTGCCGTGCACACAACTTGTACTGTTCCGGAAGATCATGAAA 358
QY 887 CTCTCTCTGAGTTGAAGCTATTCTCGAGCCTGTAAAGAACTGCATGGCCAAAGTGGAA 946
DB 357 CTCTAGCTGAGTGGAGCAATCTCTGAAGCCAGTGAAGAACTCAGACATGGCCGAGCGAA 298
QY 947 TCCATCAGAACTAA 960
DB 297 TTCAACAAAAGCTGA 284

RESULT 11

ABL06603
ID ABL06603 standard; cDNA; 885 BP.

XX
AC ABL06603;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 14291.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR P-PSDB; ABB62500.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions.

XX Claim 1; SEQ ID NO 14291; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737.

XX ABB72072). The sequence data for this patent did not form part of the

XX printed specification, but was obtained in electronic format directly

XX from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 885 BP; 205 A; 225 C; 264 G; 191 T; 0 U; 0 Other;

XX Query Match 8.4%; Score 80.2; DB 4; Length 885;

Best Local Similarity 50.5%; Pred. No. 5e-15;
Matches 249; Conservative 0; Mismatches 238; Indels 6; Gaps 2;

QY 283 TTTGATTTTCAGTGTGAGAGAGTAAGAAAGAGTATTGACGAGAGCTTGGAGAGGCTTCAG 342

DB 259 TTTGACTATTTCGGCTGACAAAGCTCGGAGAGTGTGAAGCGGAGTCTGGAGCGGCTCCAG 318

QY 343 CTTGATTATGTTGACATACATCTTCATTCGTCATGACATTTAGTTCGGGTC---TCTTGATCAG 399

DB 319 TTGGACAGGGTGGACACTACTACAGGTTTCATGACGTGGACGCGGCACCCAAATCTAGACATA 378

QY 400 ATTGTGAGTGAACAAATTCCTGCTCTTCAGAAAATCTGAAAACAGAGGGGGAAGACCGGTTTC 459

DB 379 GTGCTGAATGAGACCATACCCGTCCTCGAGGAGTACGTCCAGCGGGGAAAGGCTCGATTTC 438

QY 460 ATTGTGATCACTGCTGCTTCCGTTAGATATTTTCACTTTATGTTTCTTGATCGAGTGCCTCCA 519

DB 439 ATCGGAGTCAACCGCTACGATGTGACGCTGCTGAAGGAGTGTCCGAGCGGGGCAAGGCT 498

QY 520 GGGACTGTCGATGATATTTGTCATCTGTCATTTACGGCGGTTAATGATTCGACGTTGCTG 579

DB 499 CGCATTC---AGGTGGTGTCTCAACTATGCTCGTTACACCTTTTAGACACACACCTTGCTG 555

QY 580 GATTTACTACCTTACTTGAAGAGCAAGGTTGSGTGTGATAAGTGTCTTCTCCATTAGCA 639

DB 556 CGCTACATGAAGGACTTCCAGAAAATGGAGATGGGCGTTGCTGTGCGGCGCTCACTCA 615

QY 640 ATGGGCTCTCTTACAGAAACAAGGTCCTCTGAAATGGCAACCTGCTTCCCCTGAGCTCAAG 699

DB 616 TTGGGACTCTTGAGAAACGCTGGACCATGATCGCATCCCGGTAGTCAGGAAATCCTG 675

QY 700 TCTGCAAGCAAAAGCGCAGTTGCTCAGTGAATAAAGGGGCAAGAGATCACAAGTTA 759

DB 676 GCCGTGGCCAAACGGGGGCGGAAATCTGCCAGCAGAGAGAAACGTGGAGCTGGGAAAGCTG 735

QY 760 GCTCTGCAATACA 772

DB 736 GCCATGTACTATA 748

RESULT 12

ABL20520/C

ID ABL20520 standard; DNA; 3724 BP.

XX AC ABL20520;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 13033.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions.


```
Db      254 AGGTGGAT 262
RESULT 14
ABL06602/c
ID ABL06602 standard; cDNA; 3014 BP.
XX
XX
AC ABL06602;
XX
XX
DT 26-MAR-2002 (first entry)
XX
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 14288.
XX
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
XX
OS Drosophila melanogaster.
XX
XX
PN WO200171042-A2.
XX
XX
PD 27-SEP-2001.
XX
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
XX
PR 23-MAR-2000; 2000US-0191637P.
XX
XX
PR 11-JUL-2000; 2000US-00614150.
XX
XX
PA (PEKE ) PE CORP NY.
XX
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX
DR WPI; 2001-656860/75.
XX
XX
DR P-PSDB; ABB62499.
XX
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
XX
PS Claim 1; SEQ ID NO 14288; 21pp + Sequence Listing; English.
XX
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABLO1840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX
SQ Sequence 3014 BP; 852 A; 644 C; 610 G; 908 T; 0 U; 0 Other;
Query Match 5.9%; Score 56.6; DB 4; Length 3014;
Best Local Similarity 48.6%; Pred. No. 6.1e-07;
Matches 186; Conservative 0; Mismatches 194; Indels 3; Gaps 1;
QY 390 TCTTGATCAGATTGTGAGTGAACAATTCCTGCTCTTCAGAAACTGAAACAGAGGGGAA 449
DB 1517 TCTAGACATAGTGTGATGAGACCATACCCGCTCGAGGAGTAGCTCCAGCGGGAAA 1458
QY 450 GACCCGGTTTCATGGTATCAGTGGTCTTCGGTTAGATATTTTCACTATGTTCTTGATCG 509
DB 1457 GGTGATTCATCGGAGTCAACGGCTTACGATGGACGCTGCTGAAGGAGTGTGCGGAGCG 1398
QY 510 AGTGCTCCAGGAGCTGTCGATGATATGTCATCTGTCATCTACGCGTTAATGATTC 569
DB 1397 GGGCAAGGGTGCAC---TTCAGGTGGTGTCTCAACTATGCTCGTTACACCCCTTTTACAA 1341
QY 570 GACGTTGCTGGATTTACTACCTTACTTTGAAGAGCAAGGTGTGGGTGTGATAGTGCTTC 629
DB 1340 CACCTTGCTCGCTACATGAAGGACTTCCAGAAATGGAGGTGGGGTGTGTGTGTCGCGC 1281
```

```
QY 630 TCCATTAGCAATGGCCTCCTTACAGAACAGGTCTCTCTGAATGGACCCCTGCTTCCCC 689
DB 1280 CGCTCACTCATTTGGGACTCTTGAGAAACGCTGGACCAATCGCATCCCGGTAGTCA 1221
QY 690 TGAGCTCAAGTCTGCAACAAAGCGCAGTTGCTCACTGCAATCAAAGGGCAAGAAGAT 749
DB 1220 GGAATCTCTGGCGTGGCCAAACGGGGGGCGGAATCTGCCAGCAGAGGAACGTGGAGCT 1161
QY 750 CACAAAGTTAGCTCTGCAATACA 772
DB 1160 GGGAAAGCTGGCCATGTACTATA 1138
RESULT 15
ABL28824/c
ID ABL28824 standard; DNA; 3308 BP.
XX
XX
AC ABL28824;
XX
XX
DT 26-MAR-2002 (first entry)
XX
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 37945.
XX
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
XX
OS Drosophila melanogaster.
XX
XX
PN WO200171042-A2.
XX
XX
PD 27-SEP-2001.
XX
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
XX
PR 23-MAR-2000; 2000US-0191637P.
XX
XX
PR 11-JUL-2000; 2000US-00614150.
XX
XX
PA (PEKE ) PE CORP NY.
XX
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX
DR WPI; 2001-656860/75.
XX
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
XX
PS Claim 1; SEQ ID NO 37945; 21pp + Sequence Listing; English.
XX
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABLO1840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX
SQ Sequence 3308 BP; 898 A; 727 C; 732 G; 951 T; 0 U; 0 Other;
Query Match 5.9%; Score 56.2; DB 4; Length 3308;
Best Local Similarity 53.4%; Pred. No. 8.8e-07;
Matches 165; Conservative 0; Mismatches 138; Indels 6; Gaps 2;
QY 63 TGGTGCCTCTCGCTCGGAGTGTCTCGTCCAGTCGCGAAGATGATCGCTGCCAC 122
DB 2167 TGGTGCCACTCTCTCAAGCTCTTCCGATGACTTCGATCCGAGAGGGATCTCTAC 2108
QY 123 CGTGGCGGAGGCTTTTCGGTCTCGGTATCAACTTTCGACACCTCCCGTATTATGGAGG 182
DB 2107 GGTGCGAGGAGCCATTAGATCCGGTATCAACTACATAGACACGGCTCCCTTTTATGGCCA 2048
```

Qy	183	AACACTGTCGAGAAATGCTTGGTAAGGACTAAAGGCTTTGC-----AAGTCCTAGA	237
Db	2047	AGGCAATCGGAAGAGCTGCTTGGCCAGGCGCTCAAGGATGTGCCCGGAGGCCCTATTA	1988
Qy	238	AGTGACTACATTGTGGCTACTAAGTGTGGTAGATATAAGAGGT-TTTGATTTCAGTGC	296
Db	1987	TATAGCAACTAAAGTTGCACGTTACGAGTTGGATCCAAACAATATGTTTCGACTATACGGC	1928
Qy	297	TGAGAGAGTAAGAAAGAGTATTGACGAGAGCTTGGAGAGGCTTCAGCTTGATTATGTTGA	356
Db	1927	TGCCAAGGCTCGGAGAGTGTGAAGCGTAGTCTGGAGCTGCTCCAGTTGGACAGGGTGA	1868
Qy	357	CATACTTCA	365
Db	1867	CGTACTACA	1859

Search completed: November 13, 2005, 05:48:14
Job time : 626 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2005, 05:29:25 ; Search time 206 Seconds
(without alignments)
7625.367 Million cell updates/sec

Title: US-10-606-300-12
Perfect score: 960
Sequence: 1 atgacgaataatagagcttcg.....gtggaatccatcagaactaa 960

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	960	100.0	960	4	US-09-630-983A-12
2	43	4.5	2682	1	US-07-855-793-3
3	42.4	4.4	999	4	US-09-902-540-3489
4	42.4	4.4	17188	4	US-09-902-540-1166
5	40.4	4.2	1005	4	US-09-328-352-1953
6	40	4.2	1104	4	US-09-252-991A-10047
7	40	4.2	2514	4	US-09-252-991A-9791
8	38	4.0	1230025	4	US-09-198-452A-1
9	38	4.0	1230230	4	US-09-438-185A-1
10	37	3.9	49617	4	US-09-596-002-28
11	36.6	3.8	346112	4	US-09-949-016-13165
12	35.6	3.7	1026	4	US-09-902-540-3767
13	35.6	3.7	20303	4	US-09-902-540-1183
14	35.2	3.7	1170	3	US-09-134-001C-1021
15	35.2	3.7	3090	4	US-09-710-279-3603
16	35.2	3.7	3319	3	US-08-335-844A-15
17	35.2	3.7	3319	4	US-09-129-366-15
18	34.8	3.6	1674	4	US-09-543-681A-4029
19	34.6	3.6	1062	4	US-10-092-263-7
20	34.4	3.6	966	3	US-09-134-001C-1063
21	34.4	3.6	1621	1	US-07-937-609-22
22	34.4	3.6	1621	3	US-08-029-170-22
23	34.4	3.6	1621	4	US-09-443-745-22
24	34.2	3.6	363	4	US-09-248-796A-5576
25	34.2	3.6	1077	4	US-09-252-991A-14964
26	34.2	3.6	1155	4	US-09-252-991A-14787
27	34	3.5	1978	1	US-08-137-175A-4

28	34	3.5	1978	3	US-08-479-017-4	Sequence 4, Appli
c 29	34	3.5	9748	4	US-09-949-016-17220	Sequence 17220, A
c 30	34	3.5	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c 31	34	3.5	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c 32	33.8	3.5	832	4	US-09-621-976-2813	Sequence 2813, Ap
c 33	33.8	3.5	885	4	US-09-540-236-894	Sequence 894, App
c 34	33.8	3.5	300598	4	US-09-949-016-11868	Sequence 11868, A
c 35	33.8	3.5	302604	4	US-09-949-016-14588	Sequence 14588, A
c 36	33.8	3.5	302604	4	US-09-949-016-14589	Sequence 14589, A
c 37	33.8	3.5	308362	4	US-09-949-016-17119	Sequence 17119, A
c 38	33.6	3.5	3492	4	US-08-956-171E-34	Sequence 34, Appl
c 39	33.6	3.5	3492	4	US-08-781-986A-34	Sequence 34, Appl
c 40	33.6	3.5	7218	1	US-08-232-463-14	Sequence 14, Appl
c 41	33.6	3.5	7218	4	US-09-949-016-14679	Sequence 14679, A
c 42	33.4	3.5	5504	4	US-09-902-540-716	Sequence 716, App
c 43	33	3.4	999	4	US-09-107-532A-3237	Sequence 3237, Ap
c 44	32.8	3.4	955	4	US-09-620-312D-524	Sequence 524, App
c 45	32.8	3.4	77586	4	US-09-949-016-13220	Sequence 13220, A

ALIGNMENTS

RESULT 1
US-09-630-983A-12
; Sequence 12, Application US/09630983A
; Patent No. 6630330
; GENERAL INFORMATION:
; APPLICANT: Porro, Danilo
; APPLICANT: Sauer, Michael
; TITLE OF INVENTION: Ascorbic Acid Production from Yeast
; FILE REFERENCE: 2028.594000
; CURRENT APPLICATION NUMBER: US/09/630.983A
; CURRENT FILING DATE: 2000-08-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 12
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-630-983A-12

Query Match		100.0%	Score 960;	DB 4;	Length 960;
Best Local Similarity		100.0%	Pred. No. 3.4e-300;		
Matches 960;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	ATGACGAAATAGAGCTTCGAGCTTTGGGAAACACAGGGCTTAAGTTAGCGCCGCTTGGT	60		
Db	1	ATGACGAAATAGAGCTTTTCGAGCTTTGGGAAACACAGGGCTTAAGTTAGCGCCGCTTGGT	60		
Qy	61	TTTGGTGCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGGAAGATGATGCCGCTGCC	120		
Db	61	TTTGGTGCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGGAAGATGATGCCGCTGCC	120		
Qy	121	ACCGTGGCGAGGCTTCCGCTCGGTATCAACTTCTTCGACACCTCCCGCTTATATGGA	180		
Db	121	ACCGTGGCGAGGCTTCCGCTCGGTATCAACTTCTTCGACACCTCCCGCTTATATGGA	180		
Qy	181	GGACACCTCTGAGAAATGCTTGGTAAGGAGCTTAAGGCTTTCGAAAGTCCCTAGAGT	240		
Db	181	GGACACCTCTGAGAAATGCTTGGTAAGGAGCTTAAGGCTTTCGAAAGTCCCTAGAGT	240		
Qy	241	GACTACATTGGGCTACTAAGTGTGGTATATAAAGAGGTTTTGATTTTCAGTCTGAG	300		
Db	241	GACTACATTGGGCTACTAAGTGTGGTATATAAAGAGGTTTTGATTTTCAGTCTGAG	300		
Qy	301	AGAGTAAGAAAGATATTGACGAGCTTGGAGGCTTCAGCTTGATTTGACATA	360		
Db	301	AGAGTAAGAAAGATATTGACGAGCTTGGAGGCTTCAGCTTGATTTGACATA	360		
Qy	361	CTTCATTGCCATGACATTGAGTTCGGTCTCTTCATCAGATTGTGAGTGAACAATTCCT	420		
Db	361	CTTCATTGCCATGACATTGAGTTCGGTCTCTTCATCAGATTGTGAGTGAACAATTCCT	420		

Qy	421	GCTCTTCAGAAACTGAAACAGAGGGGAAGACCCGGTTCAATTGGTATCACTGGTCTTCGG	480
Db	421	GCTCTTCAGAAACTGAAACAGAGGGGAAGACCCGGTTCAATTGGTATCACTGGTCTTCGG	480
Qy	481	TTAGATATTTTCCACCTTATGTTCTTGATCCGAGTGCTCCAGGACTGTGCGATGTGATATG	540
Db	481	TTAGATATTTTCCACCTTATGTTCTTGATCCGAGTGCTCCAGGACTGTGCGATGTGATATG	540
Qy	541	TCNACTGTGCATTA CGGCGTTAATGATTCGACGTTGCTGGATTTATCTACTTCTTGAAG	600
Db	541	TCNACTGTGCATTA CGGCGTTAATGATTCGACGTTGCTGGATTTATCTACTTCTTGAAG	600
Qy	601	AGCAAAGGTGCGGTGTGATAAGTGCTTCTCCATTAGCAATGGCGCTCCTTTACAGAA	660
Db	601	AGCAAAGGTGCGGTGTGATAAGTGCTTCTCCATTAGCAATGGCGCTCCTTTACAGAA	660
Qy	661	GGTCCTCTGAATGCGACCTTGCTTCCCTGAGCTCAAGTCTGCAAGCAAAAGCCGAGTT	720
Db	661	GGTCCTCTGAATGCGACCTTGCTTCCCTGAGCTCAAGTCTGCAAGCAAAAGCCGAGTT	720
Qy	721	GCTCACTGCAAAATCAAGGGCAAGAAGATCAAAAGTTAGCTCTGCAATACAGTTTAGCA	780
Db	721	GCTCACTGCAAAATCAAGGGCAAGAAGATCAAAAGTTAGCTCTGCAATACAGTTTAGCA	780
Qy	781	AACAAGGAGATTCGTCGGTGTGGTTGGGATGAGCTCTGTCTCACAGGTAGAAGAAAT	840
Db	781	AACAAGGAGATTCGTCGGTGTGGTTGGGATGAGCTCTGTCTCACAGGTAGAAGAAAT	840
Qy	841	GTTCGACGAGTTACAGAGCTTCAAGAGTCTGGGGATGGATCAAGAAATCTGTCTGAGGTT	900
Db	841	GTTCGACGAGTTACAGAGCTTCAAGAGTCTGGGGATGGATCAAGAAATCTGTCTGAGGTT	900
Qy	901	GAAGCTATTCTCGAGCCTGTAAAGAAATCTGACATGGCCAAGTGGAAATCCATCAGAACTAA	960
Db	901	GAAGCTATTCTCGAGCCTGTAAAGAAATCTGACATGGCCAAGTGGAAATCCATCAGAACTAA	960

RESULT 2

US-07-855-793-3
Sequence 3, Application US/07855793
Patent No. 5217880
GENERAL INFORMATION:
APPLICANT: Masanori MITTA et al.
TITLE OF INVENTION: L-FUCOSE DEHYDROGENASE GENE,
TITLE OF INVENTION: MICROORGANISM HAVING SAID GENE AND PRODUCTION OF L-FUCOSE
TITLE OF INVENTION: DEHYDROGENASE BY THE USE OF SAID MICROORGANISM
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 500 Kb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: DisplayWrite
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/855,793
FILING DATE: 19920323
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850

; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 3489
; LENGTH: 999
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-3489

Query Match 4.4%; Score 42.4; DB 4; Length 999;
Best Local Similarity 58.9%; Pred. No. 0.006;
Matches 73; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
Qy 97 GTGCGGAGGATGATCGCGTCCACCGTGGCGGAGGCTTTCGGTCTCGGTATCAACTTC 156
Db 82 GTGGAGGAGGAGCGCGCTCGCTGCGTGGCGCGCGCTGGAGCTGGGGCATCACAC 141
Qy 157 TTCGACACCTCCCGTATTATGGAGGACACTGTCTGAGAAATGCTTGGTAAGGGACTA 216
Db 142 TTGACACCGGAGCTGTACGGCAACCGCGCGGAGGAGTGTCTGGCGGTGCCCTG 201
Qy 217 AAGG 220
Db 202 AAGG 205

RESULT 4
US-09-902-540-1166/c
; Sequence 1166, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wigand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1166
; LENGTH: 17188
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1166

Query Match 4.4%; Score 42.4; DB 4; Length 17188;
Best Local Similarity 58.9%; Pred. No. 0.036;
Matches 73; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
Qy 97 GTGCGGAGGATGATCGCGTCCACCGTGGCGGAGGCTTTCGGTCTCGGTATCAACTTC 156
Db 8987 GTGGAGGAGGAGCGCGCTCGCTGCGTGGCGCGCGCTGGAGCTGGGATCACAC 8928
Qy 157 TTCGACACCTCCCGTATTATGGAGGACACTGTCTGAGAAATGCTTGGTAAGGGACTA 216
Db 8927 TTCGACCGGAGCGTGTACGGCAACCGCGCGGAGGAGTGTCTGGCGGTGCCCTG 8868
Qy 217 AAGG 220
Db 8867 AAGG 8864

RESULT 5
US-09-328-352-1953
; Sequence 1953, Application US/09328352
; Patent No. 6562958

; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1953
; LENGTH: 1005
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1953

Query Match 4.2%; Score 40.4; DB 4; Length 1005;
Best Local Similarity 53.1%; Pred. No. 0.027;
Matches 86; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
Qy 7 AAAATAGAGCTTCAGAGCTTTGGGGAACACAGGGCTTAAGGTTAGCGCGTTTGGTTTGGT 66
Db 1 AAAATGAATATAGAACTTTAGGACAAACAGGCGGAAAAAGTATCAGCACCTTGGTTAGGT 60
Qy 67 GCCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCCGCGAAGATGATGCCGTGCCACCCGTG 126
Db 61 TGCATGGGAATGAGCTTTGCTATGCTGCTCAGACGACACAAAGCATTTGCAACTCTA 120
Qy 127 CGCGAGGCTTTCCGTCGGTATCAACTTCTTCGACACCTCC 168
Db 121 GAAAGGCTTAGACCTCGGCACTTTTGGGACACTGCC 162

RESULT 6
US-09-252-991A-10047
; Sequence 10047, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10047
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10047

Query Match 4.2%; Score 40; DB 4; Length 1104;
Best Local Similarity 54.9%; Pred. No. 0.039;
Matches 79; Conservative 0; Mismatches 65; Indels 0; Gaps 0;
Qy 54 CGTTGGTTTTGGTCCCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGCGGAAGATGATGC 113
Db 153 CATCGGTCTCGCGCGGACCGCTGGGCAACATGTTCCATCGCTCAGCGAAGAACCGC 212
Qy 114 CGTCCGACCGTGGCGGAGGCTTTCGGTCTCGGTATCAACTTCTTCGACACTCCCGTA 173
Db 213 CGACGCCACCTCAATGCCGCTGGGACGCGGCTTTCGCTATTACGAGTGTGCCCGCA 272
Qy 174 TTATGGAGGAACACTGTCTGAGAA 197
Db 273 CTACGCGCGGCGCTCGCCGAGCA 296

RESULT 7
US-09-252-991A-9791/c
; Sequence 9791, Application US/09252991A

```
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: ABRUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9791
; LENGTH: 2514
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9791

Query Match          4.2%; Score 40; DB 4; Length 2514;
Best Local Similarity 54.9%; Pred. No. 0.065;
Matches 79; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 54 CGTGTGTTTGGTGCTCTCGCTCGGAAGTGTCTTCGTCACGTCGCGAAGATGATGC 113
Db 1030 CATCGTCTCGCGGGCGACCGCTGGGCAACATGTTCCATCCGCTCAGCGAAGAACCGC 971

QY 114 CGTCGCCACCGTCGCGAGGCTTTCCGTCCTCGGTATCACTTCTTCGACACCTCCCGTA 173
Db 970 CGAAGCCACCTCAATGCCGCTGGAGCGCGCTTCGCTATTACGAGGTGCGCGCA 911

QY 174 TTATGGAGAACATGCTCTGAGAA 197
Db 910 CTACGCGCGGCGCTCGCGAGCA 887

RESULT 8
US-09-198-452A-1/c
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griflais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 5849
; SEQ ID NO 1
; LENGTH: 1230025
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(15000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (15001)..(30000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (30001)..(45000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (45001)..(60000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (60001)..(75000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (75001)..(90000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (90001)..(105000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (105001)..(120000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (120001)..(135000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (135001)..(150000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (150001)..(165000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (165001)..(180000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (180001)..(195000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (195001)..(210000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (210001)..(225000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (225001)..(240000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (240001)..(255000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (255001)..(270000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (270001)..(285000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (285001)..(300000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (300001)..(315000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (315001)..(330000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (330001)..(345000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (345001)..(360000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (360001)..(375000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (375001)..(390000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (390001)..(405000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (405001)..(420000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (420001)..(435000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (435001)..(450000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (450001)..(465000)
```


	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (465001)..(480000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (480001)..(495000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (495001)..(510000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (510001)..(525000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (525001)..(540000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (540001)..(555000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (555001)..(570000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (570001)..(585000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (585001)..(600000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (600001)..(615000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (615001)..(630000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (630001)..(645000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (645001)..(660000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (660001)..(675000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (675001)..(690000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (690001)..(705000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (705001)..(720000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (720001)..(735000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (735001)..(750000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (750001)..(765000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (765001)..(780000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (780001)..(795000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (795001)..(810000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc feature
	LOCATION: (810001)..(825000)
	OTHER INFORMATION: n=a or c or g or t

```

; NAME/KEY: misc_feature
; LOCATION: (825001)..(840000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (840001)..(855000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (855001)..(870000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (870001)..(885000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (885001)..(900000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (900001)..(915000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature

Query Match 4.0%; Score 38; DB 4; Length 1230025;
Best Local Similarity 51.1%; Pred. No. 14;
Matches 89; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

Qy 255 TACTAAGTGTGGTAGATATAAAGAGGTTTGTGATTTTCAGTGTGTGAGAGATGAGAAAGAG 314
Db 473798 TCCTGAGGGGGGAAGACTGATCTGTGTTTGGTGTGGGTAGAGATCGACGTAA 4737739

Qy 315 TATTGACGAGAGCTTGGAGAGGCTTCAGCTTGATTATGTTGACATACCTTCATTGCCATGA 374
Db 473738 ACGGAATGTTGATGGCCAGGTGTGTAGAGCGTTAATGGTTTTGCTGTTGTAACCTTCAGATAA 473679

Qy 375 CATTTGAGTTCCGGGTCTCTTGTATCAGATTTGTAGTGAACAATTTCTCTCTCTTCA 428
Db 473678 CCCTAGAGCGAGCCCTCTGAAATATTTGTGATGAAATTTGTGATCGGTTTTA 473625

```

```

RESULT 9
US-09-438-185A-1/c
; Sequence 1, Application US/09438185A
; Patent No. 6822071
; GENERAL INFORMATION:
; APPLICANT: Stephens, Richard
; APPLICANT: Mitchell, Wayne
; APPLICANT: Kalman, Sue
; APPLICANT: Davis, Ronald
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Chlamydia Pneumoniae Genome Sequence
; FILE REFERENCE: 018941-000411US
; CURRENT APPLICATION NUMBER: US/09/438,185A
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: US 60/108,279
; PRIOR FILING DATE: 1998-11-12
; PRIOR APPLICATION NUMBER: US 60/128,606
; PRIOR FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 1074
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1230230
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
US-09-438-185A-1

```

	Query Match	4.0%;	Score 38;	DB 4;	Length 1230230;
	Best Local Similarity	51.1%;	Pred. No. 14;		
	Matches 89;	Conservative 0;	Mismatches 85;	Indels 0;	Gaps 0;
Qy	255	TACTAGTGGGTAGATATAAAGAGGTTTTGATTTTCAGTGGCTGAGAGAGTAAAGAAAGAG	314		
Db	463310	TCCTGAGGGGGGAAGACTGATTGTTTGGTTCGGTGGAGATAGAGATCGCAGTAA	463251		
Qy	315	TATTGACGAGAGCTTGGAGAGGCTTCAGCTTGATTATGTTTGACATATCTTCATTCGCCATGA	374		

Db 463250 ACGGAAGTTGATGGCCAGGTGGTAGAGCGTTATGTTTGTCTGTTGTAACCTTCAGATAA 463191
QY 375 CATTGAGTTCGGGTCTCTGTGATCAGATTGTGAGTGAAGAAACAATTCCTGCTCTTCA 428
Db 463190 CCTAGGAGCGAGCCCTCCTGAAGATATTGTGAATGAAATTTGTGATGGTTTTTA 463137
RESULT 10
US-09-596-002-28
; Sequence 28, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 28
; LENGTH: 49617
; TYPE: DNA
; ORGANISM: M. catarrhalis
; NAME/KEY: misc feature
; FEATURE:
; OTHER INFORMATION: Incyte template ID No. 6632636 28
; PUBLICATION INFORMATION:
US-09-596-002-28
Query Match 3.9%; Score 37; DB 4; Length 49617;
Best Local Similarity 48.8%; Pred. No. 4;
Matches 100; Conservative 0; Mismatches 105; Indels 0; Gaps 0;
QY 411 AACAAATCTGCTCTTCAGAACTGAAACAGAGGGGAGAGCCCGGTTTCATTGGTATCAC 470
Db 5332 AGCAACCAAGTCGTGAGATTATGACACACGCCCGCAGGTGTCGATTGCCAGCGA 5391
QY 471 TGGCTTCGGTTAGATATTTTACATTATGTTCTTCATCGAGTGCCTCCAGGAGCTGTGGA 530
Db 5392 TGATGATTTATCTGATATTTTATCGGTGGTGTCTGAAACAGAGCAATTCGCTATCCTGT 5451
QY 531 TGTGATATTTGCATCTGCTCAATTCAGGGGTTAAATGATTCGAGGTTGCTGGATTACTACC 590
Db 5452 TTTTCACAGTTAGATGATGCTGTGTTGGGATTTTGTGATTAAGGACTTAATACC 5511
QY 591 TTACTTGAAGAGCAAGGTGTGGGT 615
Db 5512 ATACCTAAAGGCAAAAGCTGACGGT 5536

RESULT 11
US-09-949-016-13165/c
; Sequence 13165, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13165
; LENGTH: 346112
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(346112)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13165
Query Match 3.8%; Score 36.6; DB 4; Length 346112;
Best Local Similarity 47.6%; Pred. No. 18;
Matches 108; Conservative 0; Mismatches 119; Indels 0; Gaps 0;
QY 278 AAGTTTGTGATTTTCAGTCTGAGAGAGTAAGAAAGAGATTGACGAGAGCTTTGGAGAGGC 337
Db 344053 AAGTTTTTGGCATCCACAGCCAGGCGGTGAGCAGCAGTCAGCATTTGCTTTTGTACTCT 343994
QY 338 TTCAGCTTGATATTGTCACATCTTCAATTCGCCATGACATTTGAGTTCGGGTCTCTTGATC 397
Db 343993 TGCTGGAGGCTGGTCATGACATCTGCTGGGCCAGTTTCATCTTTGATGAGCTCACCC 343934
QY 398 AGATTGTGAGTGAACAATTCCTGCTCTTCAGAAACTGAAAACAAGAGGGGAAGACCCGGT 457
Db 343933 AGTCAGAGTTCAATAGCTTCTGTCATCTCAATCTGAAAGACAAGAGATAGGTACGA 343874
QY 458 TCATTGGTATCATCTGGTCTTCCGTTAGATATTTTCACTTATGTTCTT 504
Db 343873 GAACGTGTTTCAGTGATTTTTTTTTTTCTTTCTTTTAGAGAGATGTTCTT 343827
RESULT 12
US-09-902-540-3767
; Sequence 3767, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 3767
; LENGTH: 1026
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-3767
Query Match 3.7%; Score 35.6; DB 4; Length 1026;
Best Local Similarity 57.0%; Pred. No. 0.98;
Matches 65; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY 102 CGAAGATGATCCCGTCGCCACCGTCGCGAGGCTTTTCGGTATCAACTTCTTCGA 161
Db 96 CGACGACGAGTCGGTGGGACCATCCACGCGCGCTCGACCTGGGCATCACGCTCATCGA 155
QY 162 CACCTCCCGTATTATGAGGAACAACCTGTCTGAGAAAATGCTTGTGTAAGGACT 215
Db 156 CACGCGCGCGCATACGCGCTTCGCTCACTCCGAGAAGGTGCTGGGAGGGCGCT 209
RESULT 13
US-09-902-540-1183/c
; Sequence 1183, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.

```
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1183
; LENGTH: 20303
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1183

Query Match          3.7%; Score 35.6; DB 4; Length 20303;
Best Local Similarity 57.0%; Pred. No. 6.4;
Matches 65; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 102 CGAAGATGATCCGTCGCCACCGTCGCGAGGCTTTCGTCGCGTATCAACTTCTTCGA 161
Db 7463 CGAGCAGAGTCGGTGGCGACCATCCACGCGCGCTCGACCTGGGCATCAGCTCATCGA 7404

Qy 162 CACCTCCCGGTATATGAGAGAACACTGTCGAGAAATGCTTGGTAAGGACT 215
Db 7403 CACCGCGCGGCATACGCGCTTCGCTCACTCCGAGAAAGTCTGGCGAGGCGCT 7350

RESULT 14
US-09-134-001C-1021/c
; Sequence 1021, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1021
; LENGTH: 1170
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1021

Query Match          3.7%; Score 35.2; DB 3; Length 1170;
Best Local Similarity 48.1%; Pred. No. 1.4;
Matches 100; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

Qy 344 TTGATTATGTTGACATCTTCATGCCAAGACATTTAGTTCGGGTCTCTTGATCAGATTG 403
Db 903 TTCAATAAAGTCACCTAAATCTTTTGGATATATTGAACATGTTTCATCTGAATCAACTTG 844

Qy 404 TGAGTGAACAATCTCTCTCTTCAGAACTGAAACAGAGGGGAGACCCGGTTCATTG 463
Db 843 ATCTACTGAAAGACATCTTGTATCTGAAAGCAGAGTCATAAAATGCATGACCATTATGATG 784

Qy 464 GTATCACTGGTCTTCGGTTAGATATTTTCACTATGTTTGTATCGAGTGCCTCCAGGGA 523
Db 783 TTTTACCTTTTCAGCTGTGTATATGATGATTTAGTCTTTCCTTGAGCAATATCATCGTAAT 724

Qy 524 CTGTGATGTGATATTTGTCATCTGCA 551
Db 723 GTCTCGACCGGCCAATTTCTATAGATTCA 696

RESULT 15
US-09-279-279-3603
```

```
; Sequence 3603, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3603
; LENGTH: 3090
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3603

Query Match          3.7%; Score 35.2; DB 4; Length 3090;
Best Local Similarity 48.1%; Pred. No. 2.6;
Matches 100; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

Qy 344 TTGATTATGTTGACATCTTCATTTGCCATGACATTTGAGTTGCGGTCTCTTGATCAGATTG 403
Db 1684 TTCAATAAAGTCACCTAAATCTTTTGGATATATTGAACATGTTTCATCTGAATCAACTTG 1743

Qy 404 TGAGTGAACAATCTCTCTCTTCAGAACTGAAACAGAGGGGAGACCCGGTTCATTG 463
Db 1744 ATCTACTGAAAGACATCTTGTATCTGAAAGCAGAGTCATAAAATGCATGACCATTATGATG 1803

Qy 464 GTATCACTGGTCTTCGGTTAGATATTTTCACTTATGTTTCTTGATCGAGTGCCTCCAGGGA 523
Db 1804 TTTTACCTTTTCAGCTGTGTATATGATGATTTAGTCTTTCCTTGAGCAATATCATCGTAAT 1863

Qy 524 CTGTGATGTGATATTTGTCATCTGCA 551
Db 1864 GTCTCGACCGGCCAATTTCTATAGATTCA 1891

Search completed: November 13, 2005, 08:07:16
Job time : 217 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2005, 05:37:56 ; Search time 943 Seconds
(without alignments)
8418.904*Million cell updates/sec

Title: US-10-606-300-12
Perfect score: 960
Sequence: 1 atgacgaataagagcttcg.....gtggaatccatcagaactaa 960

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 9794790 seqs, 4134909567 residues

Total number of hits satisfying chosen parameters: 19589580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*
23: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
24: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
25: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq.*
26: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
27: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
28: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	960	100.0	960	19	US-10-240-136A-4
2	960	100.0	960	22	US-10-606-300-12
3	522.6	54.4	1361	19	US-10-424-599-119463
4	454	47.3	1616	20	US-10-437-963-11457
5	426	44.4	1715	21	US-10-425-115-95235

6	414.8	43.2	1434	20	US-10-767-701-13219	Sequence 13219, A
7	384.4	40.0	1295	21	US-10-425-115-95234	Sequence 95234, A
8	340.4	35.5	1134	19	US-10-425-114-31543	Sequence 31543, A
9	335	34.9	564	20	US-10-021-323-3555	Sequence 3555, Ap
10	224.8	23.4	563	20	US-10-021-323-3629	Sequence 3629, Ap
11	110	11.5	1038	26	US-11-097-143-28022	Sequence 28022, A
12	99.4	10.4	342	21	US-10-425-115-29198	Sequence 29198, A
13	98.2	10.2	1051	26	US-11-097-143-40478	Sequence 40478, A
14	95.8	10.0	570	20	US-10-021-323-2292	Sequence 2292, Ap
15	94	9.8	255	19	US-10-425-114-23348	Sequence 23348, A
16	89	9.3	484	20	US-10-021-323-2400	Sequence 2400, Ap
17	81.2	8.5	575	20	US-10-021-323-816	Sequence 816, App
18	80.2	8.4	885	26	US-11-097-143-7145	Sequence 7145, Ap
19	71.2	7.4	3724	26	US-11-097-143-28021	Sequence 28021, A
20	68.2	7.1	984	18	US-10-282-122A-25914	Sequence 25914, A
21	63.8	6.6	2016	21	US-10-425-115-151209	Sequence 151209, A
22	56.6	5.9	3014	26	US-11-097-143-7144	Sequence 7144, Ap
23	56.2	5.9	3308	26	US-11-097-143-40477	Sequence 40477, A
24	56.2	5.9	10400	26	US-11-097-143-27337	Sequence 27337, A
25	56.2	5.9	10418	26	US-11-097-143-40462	Sequence 40462, A
26	56.2	5.9	10451	26	US-11-097-143-26110	Sequence 26110, A
27	51.6	5.4	984	16	US-10-156-761-1393	Sequence 1393, Ap
28	51.6	5.4	9025608	16	US-10-156-761-1	Sequence 1, Appli
29	46.8	4.9	984	16	US-10-156-761-7263	Sequence 7263, Ap
30	46.8	4.9	9025608	16	US-10-156-761-1	Sequence 1, Appli
31	46	4.8	586	20	US-10-021-323-17546	Sequence 17546, A
32	45.8	4.8	1892	21	US-10-425-115-58372	Sequence 58372, A
33	45	4.7	912	18	US-10-282-122A-9381	Sequence 9381, Ap
34	44.2	4.6	1249	19	US-10-425-114-15371	Sequence 15371, A
35	44.2	4.6	1351	19	US-10-425-114-1870	Sequence 1870, Ap
36	44.2	4.6	1383	19	US-10-425-114-21231	Sequence 21231, A
37	43.4	4.5	1446	21	US-10-739-930-5295	Sequence 5295, Ap
38	42.8	4.5	2838	20	US-10-437-963-68518	Sequence 68518, A
39	42.4	4.4	960	18	US-10-369-493-43084	Sequence 43084, A
40	42.4	4.4	2374	19	US-10-424-599-100601	Sequence 100601, A
41	42	4.4	975	18	US-10-369-493-28562	Sequence 28562, A
42	42	4.4	1674	21	US-10-739-930-4706	Sequence 4706, Ap
43	41.8	4.4	706	21	US-10-425-115-33492	Sequence 33492, A
44	41.8	4.4	1464	20	US-10-437-963-21571	Sequence 21571, A
45	41	4.3	328	20	US-10-767-701-16000	Sequence 16000, A

ALIGNMENTS

RESULT 1
US-10-240-136A-4
; Sequence 4, Application US/10240136A
; Publication No. US20040053235A1
; GENERAL INFORMATION:
; APPLICANT: Ascorbex Limited
; TITLE OF INVENTION: Gene Sequence
; FILE REFERENCE: JPD/P100606US-PCT
; CURRENT APPLICATION NUMBER: US/10/240,136A
; CURRENT FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: GB 0007651.3
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(960)
; OTHER INFORMATION:
US-10-240-136A-4

Query Match 100.0% Score 960; DB 19; Length 960;
Best Local Similarity 100.0%; Pred. No. 1.2e-302;
Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGACGAAATAGAGCTTCGAGCTTTGGGAAACACAGGGCTTAAGGTTAGCGCGTGGT 60
Db 1 ATGACGAAATAGAGCTTCGAGCTTTGGGAAACACAGGGCTTAAGGTTAGCGCGTGGT 60
QY 61 TTTGGTGCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGGAAGATGATGCCGTCGCC 120
Db 61 TTTGGTGCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGGAAGATGATGCCGTCGCC 120
QY 121 ACCGTGCGAGGCTTTCCGCTCGGTATCAACTCTTCGACACCTCCCGGTATTATGGA 180
Db 121 ACCGTGCGAGGCTTTCCGCTCGGTATCAACTCTTCGACACCTCCCGGTATTATGGA 180
QY 181 GGAACACTGTCTGAGAAATGCTTCGGTAAGGACTAAAGGCTTTGCAAGTCCCTAGAA 240
Db 181 GGAACACTGTCTGAGAAATGCTTCGGTAAGGACTAAAGGCTTTGCAAGTCCCTAGAA 240
QY 241 GACTACATTTGGCTTACTAAGTGTGCTAGATATAAGAAAGGTTTGTGATTCAGTCTGAG 300
Db 241 GACTACATTTGGCTTACTAAGTGTGCTAGATATAAGAAAGGTTTGTGATTCAGTCTGAG 300
QY 301 AGAGTAAGAAAGAGTATTGACGAGAGCTTTGGAGAGGCTTCAGCTTTGATTATGTTGACATA 360
Db 301 AGAGTAAGAAAGAGTATTGACGAGAGCTTTGGAGAGGCTTCAGCTTTGATTATGTTGACATA 360
QY 361 CTTCAATGTCATGCAATTCAGTTCGGGTCTCTTGATCAGATTGTGAGTGAACAATTCCT 420
Db 361 CTTCAATGTCATGCAATTCAGTTCGGGTCTCTTGATCAGATTGTGAGTGAACAATTCCT 420
QY 421 GCTCTTCAGAACTCAAAACAGAGGGGAAGACCCGGTTCATTGGTATCACTGGTCTCCG 480
Db 421 GCTCTTCAGAACTCAAAACAGAGGGGAAGACCCGGTTCATTGGTATCACTGGTCTCCG 480
QY 481 TTAGATATTTTCACTTATGTTCTTGATCGAGTGCCTCCAGGAGCTGTCGATGTGATATTG 540
Db 481 TTAGATATTTTCACTTATGTTCTTGATCGAGTGCCTCCAGGAGCTGTCGATGTGATATTG 540
QY 541 TCATACCTGTCATTCAGGCTTAATGATTCGAGTGCCTCCAGGAGCTTACTACCTTACTTGAAG 600
Db 541 TCATACCTGTCATTCAGGCTTAATGATTCGAGTGCCTCCAGGAGCTTACTACCTTACTTGAAG 600
QY 601 AGCAAGGTTGGGTGTGATAGTCTTCTCCATTAGCAATGGGCTCTCTACAGAACAA 660
Db 601 AGCAAGGTTGGGTGTGATAGTCTTCTCCATTAGCAATGGGCTCTCTACAGAACAA 660
QY 661 GGTCTCTGTAATGCGACCTCTTCCCTGAGCTCAAGTCTGCAAGCAAAAGCCGAGTT 720
Db 661 GGTCTCTGTAATGCGACCTCTTCCCTGAGCTCAAGTCTGCAAGCAAAAGCCGAGTT 720
QY 721 GCTCACTGCAATCAAAAGGCAAGAGATCAAAAGTTAGCTCTGCAATACAGTTTAGCA 780
Db 721 GCTCACTGCAATCAAAAGGCAAGAGATCAAAAGTTAGCTCTGCAATACAGTTTAGCA 780
QY 781 AACAAAGGATTTCTGTCGGTGTGTTGGGATGAGCTCTGTCTACAGTAGAAGAAAT 840
Db 781 AACAAAGGATTTCTGTCGGTGTGTTGGGATGAGCTCTGTCTACAGTAGAAGAAAT 840
QY 841 GTTGACGAGTTACAGAGCTTGAAAGTCTGGGATGGATCAAGAACTCTGTCTGAGTT 900
Db 841 GTTGACGAGTTACAGAGCTTGAAAGTCTGGGATGGATCAAGAACTCTGTCTGAGTT 900
QY 901 GAAGCTATTCGAGCCTGTAAGATCTGACATGCCCAAGTGGATCCATCAGAACTAA 960
Db 901 GAAGCTATTCGAGCCTGTAAGATCTGACATGCCCAAGTGGATCCATCAGAACTAA 960
```

RESULT 2

US-10-606-300-12

```
; Sequence 12, Application US/10606300
; Publication NO. US20050019879A1
; GENERAL INFORMATION:
; APPLICANT: Porro, Danilo
; APPLICANT: Sauer, Michael
; TITLE OF INVENTION: Ascorbic Acid Production from Yeast
```

```
; FILE REFERENCE: 2028.594000
; CURRENT APPLICATION NUMBER: US/10/606.300
; CURRENT FILING DATE: 2003-06-25
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-606-300-12
```

Query Match 100.0%; Score 960; DB 22; Length 960;

Best Local Similarity 100.0%; Pred. No. 1.2e-302; Mismatches 0; Gaps 0;

Matches 960; Conservative 0; Indels 0;

```
QY 1 ATGACGAAATAGAGCTTCGAGCTTTGGGAAACACAGGGCTTAAGGTTAGCGCGTGGT 60
Db 1 ATGACGAAATAGAGCTTCGAGCTTTGGGAAACACAGGGCTTAAGGTTAGCGCGTGGT 60
QY 61 TTTGGTGCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGGAAGATGATGCCGTCGCC 120
Db 61 TTTGGTGCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGGAAGATGATGCCGTCGCC 120
QY 121 ACCGTGCGAGGCTTTCCGCTCGGTATCAACTCTTCGACACCTCCCGGTATTATGGA 180
Db 121 ACCGTGCGAGGCTTTCCGCTCGGTATCAACTCTTCGACACCTCCCGGTATTATGGA 180
QY 181 GGAACACTGTCTGAGAAATGCTTCGGTAAGGACTAAAGGCTTTGCAAGTCCCTAGAA 240
Db 181 GGAACACTGTCTGAGAAATGCTTCGGTAAGGACTAAAGGCTTTGCAAGTCCCTAGAA 240
QY 241 GACTACATTTGGCTTACTAAGTGTGCTAGATATAAGAAAGGTTTGTGATTCAGTCTGAG 300
Db 241 GACTACATTTGGCTTACTAAGTGTGCTAGATATAAGAAAGGTTTGTGATTCAGTCTGAG 300
QY 301 AGAGTAAGAAAGAGTATTGACGAGAGCTTTGGAGAGGCTTCAGCTTTGATTATGTTGACATA 360
Db 301 AGAGTAAGAAAGAGTATTGACGAGAGCTTTGGAGAGGCTTCAGCTTTGATTATGTTGACATA 360
QY 361 CTTCAATGTCATGCAATTCAGTTCGGGTCTCTTGATCAGATTGTGAGTGAACAATTCCT 420
Db 361 CTTCAATGTCATGCAATTCAGTTCGGGTCTCTTGATCAGATTGTGAGTGAACAATTCCT 420
QY 421 GCTCTTCAGAACTCAAAACAGAGGGGAAGACCCGGTTCATTGGTATCACTGGTCTCCG 480
Db 421 GCTCTTCAGAACTCAAAACAGAGGGGAAGACCCGGTTCATTGGTATCACTGGTCTCCG 480
QY 481 TTAGATATTTTCACTTATGTTCTTGATCGAGTGCCTCCAGGAGCTGTCGATGTGATATTG 540
Db 481 TTAGATATTTTCACTTATGTTCTTGATCGAGTGCCTCCAGGAGCTGTCGATGTGATATTG 540
QY 541 TCATACCTGTCATTCAGGCTTAATGATTCGAGTGCCTCCAGGAGCTTACTACCTTACTTGAAG 600
Db 541 TCATACCTGTCATTCAGGCTTAATGATTCGAGTGCCTCCAGGAGCTTACTACCTTACTTGAAG 600
QY 601 AGCAAGGTTGGGTGTGATAGTCTTCTCCATTAGCAATGGGCTCTCTACAGAACAA 660
Db 601 AGCAAGGTTGGGTGTGATAGTCTTCTCCATTAGCAATGGGCTCTCTACAGAACAA 660
QY 661 GGTCTCTGTAATGCGACCTCTTCCCTGAGCTCAAGTCTGCAAGCAAAAGCCGAGTT 720
Db 661 GGTCTCTGTAATGCGACCTCTTCCCTGAGCTCAAGTCTGCAAGCAAAAGCCGAGTT 720
QY 721 GCTCACTGCAATCAAAAGGCAAGAGATCAAAAGTTAGCTCTGCAATACAGTTTAGCA 780
Db 721 GCTCACTGCAATCAAAAGGCAAGAGATCAAAAGTTAGCTCTGCAATACAGTTTAGCA 780
QY 781 AACAAAGGATTTCTGTCGGTGTGTTGGGATGAGCTCTGTCTACAGTAGAAGAAAT 840
Db 781 AACAAAGGATTTCTGTCGGTGTGTTGGGATGAGCTCTGTCTACAGTAGAAGAAAT 840
QY 841 GTTGACGAGTTACAGAGCTTGAAAGTCTGGGATGGATCAAGAACTCTGTCTGAGTT 900
Db 841 GTTGACGAGTTACAGAGCTTGAAAGTCTGGGATGGATCAAGAACTCTGTCTGAGTT 900
```

Db 841 GTTCAGCAGTTACAGAGCTTGAAGTCTGGGATGGATCAAGAAACTCTGTCTGAGTT 900
Qy 901 GAAGCTATTCTCAGGCTTGAAGAACTGACATGCCAAGTGGAAATCCATCAGAACTAA 960
Db 901 GAAGCTATTCTCAGGCTTGAAGAACTGACATGCCAAGTGGAAATCCATCAGAACTAA 960

RESULT 3

US-10-424-599-119463
; Sequence 119463, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 119463
; LENGTH: 1361
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_78884C.1
US-10-424-599-119463

Query Match 54.4%; Score 522.6; DB 19; Length 1361;
Best Local Similarity 71.8%; Pred. No. 2.2e-159;
Matches 684; Conservative 0; Mismatches 269; Indels 0; Gaps 0;

Qy 8 AATAGAGCTTCAGCTTTGGGAAACACAGGCTTAAGTTAGCCCGTGTGTTTGTG 67
Db 56 AATAGAGCTTCAGCTTTGGGAAACACAGGCTTAAGTTAGCCCGTGTGTTTGTG 115
Qy 68 CCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCCGAGATCATGCCGTCACCGTGC 127
Db 116 CTCTCCACTGGCAATGCTTCGGGATGTTTCCGAGMACAGCCAAAGCTTCGGTTC 175
Qy 128 GCGAGCTTTCCGTTCCGTTATCAATCTTCCAGACCTCCCGTATTATGGAGAACAC 187
Db 176 GCTCGCTTTCAATCCGCACTAATTTCTCGACACTCTCCGCTAGCGAGGACAC 235
Qy 188 TGTCTGAGAAATGCTTGTGTAAGGACTAAAGGCTTTGCAAGTCCCTAGAGTACTACA 247
Db 236 TGTCCGAAAAAGTTCTGGGAAAGGCACCTGAAGGCTCTGGAGCTCCGAGGAACCTGTACG 295
Qy 248 TTGTGCTACTAAGTGTGTAGATATAAAGAGGTTTGTATTCAGTGTGAGAGAGTAA 307
Db 296 TTGTGCAACGAAGTCCGACCGGTACAGGAGGTTTCATTTACGCGGAGAGAGTGA 355
Qy 308 GAAAGAGATTTCAGCAGAGCTTTGGAGAGCTTCAGCTTGATTATGTTGACATCTCAT 367
Db 356 CGAGAAGCATTCAAGAGAGCTTTGGAAGGTTTCAGCTTGACTACGTCGACATTTCCCAAT 415
Qy 368 GCAATGACATTGAGTTCCGGTCTCTTGATCAGATTGTGAGTGAACAAATTCCTGCTCTTC 427
Db 416 GCCAGACATCAGTTCCGCTCTTTAGACAGGTTGTGAATGAGACGATTCACAGCGCTTG 475
Qy 428 AGAAATGAAACAGAGGGGAAGACCCGGTTCAATTTGTTATCACTGTCTTCGGTTAGATA 487
Db 476 TGAAGCTGAAGAGCAGAGGAAGGCGGTTTCATTTGGAATAACAGGGCTTCTCTGGGA 535
Qy 488 TTTTCACTTATGTTCTTGATCGAGTCCCTCCAGGGACTGTGATGTGATATTTGTCATACT 547
Db 536 TTTTTCAGTTACGTGCTTGATAGGTTTCGCTCGGAGCGCTGATGTTGTGCTTTTCATATT 595
Qy 548 GTCATTACGGCTTAATGATTCGAGTTGCTCGATTTACTACTTACTTGAAGACAAAG 607
Db 596 GCCATTACTGTGTAATGACACTAGCTTTGGGGAAATTTGGTCCCTATTTTGAAGACAAAG 655

Qy 608 GTGTGGGTGTGATTAAGTGTCTTCCATTAGCAATGGGCTCTCTTACAGAACAAAGTCTCTC 667
Db 656 GGGTTGGCATTAATCAATGCTTCTCCCTGTCTATGGGCTTCTCACCGAGTCTGGCCGCG 715
Qy 668 CTGAATGGCACCCCTGCTTCCCTGAGCTCAAGCTGTGCAAGCAAAAGCCGAGTGTCTACT 727
Db 716 CTGAATGGCATCCCGCATCACTGGAACTCAAGTCTGCAATGTCAAGCTCTGCTACCCATT 775
Qy 728 GCAATCAAGGGCAAGAGATCAAAAGTTAGCTCTGCAATACAGTTTGTAGCAAAACAGG 787
Db 776 GTAAAGAAAAGGAAAACATTTCAAAGTTAGATTCAGTACAGCTTGTTAATAAGG 835
Qy 788 AGATTTTCGTCGGTGTGTTGGTGGATGAGCTCTGTCTCAAGGTAGAAGAAATGTTGAG 847
Db 836 AATACACATCAGTCTGTTAGCATGAAGTCTGTTGAACAGGTGGAGAAAATGTTGCTG 895
Qy 848 CAGTTACAGAGCTTGAAGTCTGGGATGGATCAAGAAACTCTGTCTGAGGTTGAGCTA 907
Db 896 CCGCAAGAGAACTTGCACCTTCTGGAATTGATGAAGAGACTCTGTCTGAAGTCTGAAGCTA 955
Qy 908 TTCTCGAGCTGTAAAGAACTCTGACATGGCCCAAGTGGAAATCCATCAGAACTAA 960
Db 956 TTCTGAAGCTGTATAAAACCAGTCTGTGGCCCAAGTGGAAATCCAGCAGAGCTGA 1008

RESULT 4

US-10-437-963-11457
; Sequence 11457, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 11457
; LENGTH: 1616
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_17680C.1
US-10-437-963-11457

Query Match 47.3%; Score 454; DB 20; Length 1616;
Best Local Similarity 67.1%; Pred. No. 7.1e-137;
Matches 643; Conservative 0; Mismatches 315; Indels 0; Gaps 0;

Qy 3 GACGAAATAGAGCTTCGAGCTTTGGGAAACACAGGCTTAAGTTAGCGCGTGTGTTT 62
Db 341 GAGCGACATGAGCTCCGCGAGCTCGGCGCACCGGCTCCGCGTCAGCCCGTTCGGCTT 400
Qy 63 TGGTGCCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCCGCCGAAGATGATGCCGTCGCCAC 122
Db 401 CGCGCTCTCCCTCTCGGCCATGCTTCGGCGAGCTCCCGCGAGCTCGCCCGCGCGCG 460
Qy 123 CGTGGGAGAGCTTTCCGCTCTCGGTATCAACTTCTTCGACACCTCCCGCTATTATGGAGG 182
Db 461 CGTCCGCGCGCCCTCGACCTCGGACATCACTTCTTCGACACCTCCCGCTACTACGGGG 520
Qy 183 AACACTGTCTGAGAAATGCTTGGTGAAGGACTAAAGGCTTTGCAAGTCCCTAGAAGTGA 242
Db 521 CACGGTGTGCGAGTCGGTGTCTCGGGAGCTGCTCCGCGCGCGGCGGTGCGCGGACCG 580

APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
FILE REFERENCE: 38-21(53535)B
CURRENT APPLICATION NUMBER: US/10/767,701
CURRENT FILING DATE: 2004-01-29
NUMBER OF SEQ ID NOS: 63128
SEQ ID NO 13219
LENGTH: 1434
TIPS: DNA
ORGANISM: Sorghum bicolor
FEATURE:
OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS59142_1
US-10-767-701-13219

Query Match 43.2%; Score 414.8; DB 20; Length 1434;
Best Local Similarity 65.4%; Pred. No. 4.5e-124;
Matches 624; Conservative 0; Mismatches 327; Indels 3; Gaps 1;
QY 10 ATAGAGCTTCGAGCTTTGGGGAACACAGGCTTAAGGTTAGCGCGTGTGTTTGGTGCC 69
DB 55 ATGAGCTCCGCGAGCTGGCGGTACGGCCCTCCGAGTCAGCGCTCGGCTTTGGCGCC 114
QY 70 TCTCGCTCGGAAGTGTCTTCGGTCCAGTCGCGGAGATGATGCCGTGCGCCACCGTGGCC 129
DB 115 TCCCGCTCGGAAGTGTTCGGGAGCTGCGCGGATACCGCGCGCGCGCGCTCGC 174
QY 130 GAGGCTTCCGCTCGGTATCACTTCTTCGACACTCCCGGTATTAATGAGGAACACTG 189
DB 175 CGCGGCTCGACCTCGCATCACTTCTTCGACACTCCCGGTACTACGGGGCACGGTC 234
QY 190 TCTGAGAAATCTTGGTAAGGACTTAAGGCTTTGCAAGTCCCTAGAGTCACTACATT 249
DB 235 TCGAGTCACTCTCGGCGATTGCTCCGCGCACCGCGCGTTCGCGGAGCGAGTCGTC 294
QY 250 GTGGCTACTAAGTGTGTAGATATAA---GAAGGTTTGTGATTCAGTGTCTGAGAGATA 306
DB 295 GTCCGACCAAGTGGCGCGCTACAAAGACGAGGTTTCGACTTCTCGCGGACCGTGTG 354
QY 307 AGAAGAGTATGACAGAGCTTGGAGAGGCTTCAGTGTATTAATGATGATATTCAT 366
DB 355 ACGCGCAGCATAGACAGAGGCTCGCGCGCTGGGGCTGGACTAGCTTGAATCTCCAC 414
QY 367 TGCCATGACATGAGTTCGGGTCTCTGATCAGATGTGAGTGAACAAATTCCTGCTTT 426
DB 415 GCCCAGCATGAGTTCAACCATCTCGACCAAGATTGTAATGAGACAAATTCGCGCTC 474
QY 427 CAGAACTGAAACAGAGGGGAGAGACCGGTTTCAATGTTATCCTGTTTCCGTTAGAT 486
DB 475 CAGAGATTAGGAGAGTGGGAGGACGTTTCATTTGGCATCACCGGCTGCGCTCAGC 534
QY 487 ATTTTCACTATGTTTGTATCGAGTGCTCCAGGAGTGTGATGATATGTCATAC 546
DB 535 ATCTACCTCTATGCTCTGACAGGCTAGCACCGGCTCGGTGAGCTGATCTATCTTAC 594
QY 547 TGCTATACCGGTTAATGATTGAGCTTGTGATTTACTACTTACTTACTGAGGACAA 606
DB 595 TGCCACTACGGATCAATGACACCTCCCTTGTGATCTGCTCCCTACTTGAAGAGCAA 654
QY 607 GGTGTGGGTGTGATAAGTCTCTCCATTAAGCAATGGGCTCTTTCAGAAACAGGTCCT 666
DB 655 GGTGTGGGTGTATCAGTGTCTGCGCATCTCTCCATGGGTCTTTTAAAGATATGGGCA 714
QY 667 CTGTAATGACACCTGCTTCCCTGAGTCTCAAGTCTGCAAGCAAGCCGAGTGTCTCAC 726
DB 715 CGGAGTGGCACCTGACCAAGAACTTAACTAGTCAGCATGACAGGCTGACGAGATCAC 774
QY 727 TCAATCAAGGGCAAGAGATCAAAAGTTAGTCTGCAATACAGTTTAGCAACAG 786
DB 775 TGTAAGAGAGGGGAGAGCAATACCAAGCTAGCTATGCAATGATAGCTTGTGATTAAT 834

QY 787 GAGATTTCTGTCGGTGTGTTGGATGAGCTCTGTCTCACAGGTAGAGAAAATGTTGCA 846
DB 835 GAAATCTCGACAAATCTTGTGTAATGAATCTTTCAGAACAGGTGGAGGAAATGTGCT 894
QY 847 GCAGTTACAGAGCTTGAAGTCTCGGATGATCAAGAACTCTGTCTGAGGTTGAAGCT 906
DB 895 GCTGCACTGGAGTTGTCAACATCAGGCATTCATGAAGAACTTCTGCGTGAAGTTGAAGCA 954
QY 907 ATTCTCGAGCTGTAAAGAAATCTGACATGCGCAAGTGAATCCATCAGAACTAA 960
DB 955 ATTCTGAGCCTGTGAAGAACCTGACTTGGCTAGCGGCATCCAAACAGTCTGA 1008
RESULT 7
US-10-425-115-95234
Sequence 95234, Application US/10425115
Publication No. US20040214272A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants
FILE REFERENCE: 38-21(53222)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 95234
LENGTH: 1295
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(1295)
OTHER INFORMATION: unsure at all n locations
FEATURE:
OTHER INFORMATION: Clone ID: MNT4577_186845C.1
US-10-425-115-95234

Query Match 40.0%; Score 384.4; DB 21; Length 1295;
Best Local Similarity 64.0%; Pred. No. 3.7e-114;
Matches 630; Conservative 0; Mismatches 321; Indels 33; Gaps 2;
QY 10 ATAGAGCTTCGAGCTTTGGGGAACACAGGCTTAAGGTTAGCGCGTGTGTTTGGTGCC 69
DB 1 ATGAGCTCCCGAGCTGGCGGTACGGGTCTCGCGTCAGCGCGCTTGGTGCC 60
QY 70 TCTCGCTCGGAAGTGTCTTCGGTCCAGTCGCGGAGATGATGCCGTGCGCCACCGTGGC 129
DB 61 TCTCGCTTGGCAAGTTCGCGGAGCTGCGCGGACACCGCGCGCGCGCTCGC 120
QY 130 GAGGCTTCCGCTCGGTATCAACTTCTTGACACCTCCCGGTATTAATGAGGAACTG 189
DB 121 CGCGCTCGACCTCGCATCACTTCTCGATCACTCACCGTACTACGGTGGCAGATA 180
QY 190 TCTCAGAAATGCTTGTGAAGGACTAAAGCTTTGCAAGTCCCTAGAAGTACTACATT 249
DB 181 TCGAGTCACTCTCGCGGATTCCTCGGCAACCGCGCGCTTCGCGGAGCGAGTCTC 240
QY 250 GTGCTACTAAGTGTGTAGATATAA---GAAGGTTTGTGATTTTCACTGTGAGAGATA 306
DB 241 GTCGCCCAAGTTCGCGCGCTACAAAGAGAGGTTTCGACTTCTCCGCCAACCGTGTG 300
QY 307 AGAAGAGTATTGACGAGCTTGGAGAGCTTCAGCTTGTATGTTGATGATCACTTCAAT 366
DB 301 ACACGACATAGACGAGAGCTCGCGGCTGGGCTGGAGTCTGCTGATCTCTCCAC 360
QY 367 TGCCATGACATGAGTTCGGGTCTTGTGATGATGATGAGTGAACAAATTCCTGCTCTT 426
DB 361 GCTCATGACATGAGTTCACCCATCTCGAACAGATTGTGAATGAGACAAATTCGCGACTC 420
QY 427 CAGAACTGAAACAGAGGGGAGAGACCGGTTTCAATGTTATCACTGTGTTCTTCCGTTAGAT 486

Db 421 CAGAAGATCAAGAGAAATGGGAAGGACCGTTTCATTGGCATCACCGGTTGCCTCTCAGC 480
Qy 487 ATTTTCACATATGTTCTTGATCGAGTGCCTCCAGGAGCTGTGATGTGATATTTGCATAC 546
Db 481 ATCTACCTTATGTTCTCGACCGTGTAGACACAGGCTCGGTGAGCTGATTTCTATCTTAT 540
Qy 547 TGTCAATACGGGTTAATGATTCGACGTTTCTGGATTTACTACCTTACTTTGAAGACAAA 606
Db 541 TGTCACTACGGGATCAATGACACCTCCCTTGTGTGATCTGCTCCCTTACTTTGAAGACAAA 600
Qy 607 GGTGTTGGGTGTGATAGTGTCTTCCATTAGCAATGGGCTCTTACAGAAACAAGTCTCT 666
Db 601 GGTGTTGGGTTATCATAGTGTCTTGGCCCTCTCCATGGGTCTTTTAAACAGATAATGGGCA 660
Qy 667 CTGTAATGCGACCTGCTTCCCTGAGCT-----C 696
Db 661 CCGAGTGGACCCCTGCACAGAGAACTTAAGATGCCCTGAATTTCTACTAGAAAATGNA 720
Qy 697 AAGTCTGCAAGCAAGCCGAGTTGCTCTCACTGCAATCAAGGGCAAGAGATCAAAAG 756
Db 721 AAGTCAGCATGCAAGGCTCAGCAGATCACTGTAGAAAGAGGGGAAAGCATTACGAAG 780
Qy 757 TTAGCTCTCAATACAGTTTACCAACAGGAGATTTCTGCTGTTGTTGGATGAGC 816
Db 781 CTAGCTATCAGTACAGCTTAAATGAACAATGAATCTCGACAGTTCTTGTGGAATGAAC 840
Qy 817 TCTGCTCTCAGGTAGCAAAATGTTGCAGCAGTTTACAGAGCTTGAAGATCTGGGATG 876
Db 841 TCTTTGGAACAGGTGAGAGAAATGTTGGCGCTGCACTGGAGTTGTCAACATCAGGCATC 900
Qy 877 GATCAAGAAACTTGTCTCAGGTTGAAGCTATTTCTGAGCCCTGTAAAGAAATCTGACATGG 936
Db 901 GATGATGAATATGCTGAGTTGAAGCAATTTCTGAGCCCTGTGAAGAAATCTGACATGG 960
Qy 937 CCAAGTGGAAATCCATCAGAACTAA 960
Db 961 CCTAGCGGCATCCAAACAGCCTGA 984

RESULT 8

US-10-425-114-31543
; Sequence 31543, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:

; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT FILING DATE: 2003-04-28
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 31543
; LENGTH: 1134
; TYPE: DNA
; ORGANISM: Zea mays

; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73186H07_FLI

US-10-425-114-31543

Query Match 35.5%; Score 340.4; DB 19; Length 1134;
Best Local Similarity 68.5%; Pred. No. 8.7e-100;
Matches 470; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

Qy 275 AAGAAGTTTTGATTTTCAGTGTGAGAGAGTAAAGAGATTTGACGAGCTTGGAGA 334
Db 7 ACGAGGTTTCGATTTCTCGCCCAACCGTGTGACACGACGATAGACGAGAGCTCGCCC 66

Qy 335 GGCTTCAGCTTGATTTATGTTGACATATCTTCAITTGCCATGACATTGAGTTTCGGTCTCTTG 394
Db 67 GACTGGGCTGGACCTACGTCGACATCTCTCCACGCTCATGACATCGAGTTCACTCCATCTCG 126
Qy 395 ATCAGATTGTCAGTGAAACAATTCCTGCTCTTTCAGAAAACCTGAAAACAAGGGGAAGACCC 454
Db 127 AACAGATTGTGATGAGACAAATTCGCGACATCCAGAGATCAGGAGAAATGGGAAGGCAC 186
Qy 455 GGTTCATTGGTATCACTGGTCTTCGTTAGATATTTTCACTTATGTTTCTTATGATCGAGTGC 514
Db 187 GGTTCATTGGCATCACCGGTTGCTCTCAGCATCTACCTTATGTCCTCGACCGCTGTAG 246
Qy 515 CTCAGGAGCTGTGATGTGATATTTGCTACTGTCTATACGGCTTATGATTCGACGT 574
Db 247 CACCAAGGCTCGGTGGACGTGATTTCTTATTTGTCACCTACCGGATCAATGACACCTCCC 306
Qy 575 TGCTGGATTTACTACTTACTTGAAGCAAGGTTGGGTGTGATATAAGTGTCTTCTCCAT 634
Db 307 TGTGTGATCTGCTCCCTACTTTGAAGCAAGGTTGGGTATCAGTGTCTTCCGCC 366
Qy 635 TAGCAATGGGCTCTCTTACAGAACAGGTCCTCTGAAATGGCACCTTCTCCCTGAGC 694
Db 367 TCTCCATGGTCTTTTAAACAGATAATGGGCCACCGGAGTGGCACCTGCACCAGAAGAAC 426
Qy 695 TCAAGTCTGCAAGCAAGCCGAGTTGCTCACTGCAATCAAGGGCAAGAGATCACA 754
Db 427 TTAAGTCAGCATGCAAGGCTGCAGCAGATCACTGTAGAAAGAGGGGAAAGCAATTACGA 486
Qy 755 AGTTAGCTCTGCAATACAGATTTAGCAACAAGGAGATTTCTGCGGTGTGTTGGGATGA 814
Db 487 AGCTAGCTATGAGTACAGCTTATGAACAATGAATTCGACAGTTCTTGTGTGAATGA 546
Qy 815 GCTCTGTCTCAGGTAGAGAAAATGTTGCAGCAGTTTACAGAGCTTGAAGTCTGGGGA 874
Db 547 ACTCTTTGGAACAGGTGGAGGAGAAATGTCGGCTGCACCTGAGTTGTCAACATCAGGCA 606
Qy 875 TGGATCAAGAAACTCTGCTGAGGTTGAGCTATTTCTGAGCCTGTAAGATCTGACAT 934
Db 607 TCGATGAAGAACTTATGCGTGAAGTTGAAGCAATTTCTGAGCCTGTGAAGAAATCTGACAT 666

RESULT 9

US-10-021-323-3555
; Sequence 3555, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:

; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US/10/021,323
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255, 619
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 3555
; LENGTH: 564
; TYPE: DNA
; ORGANISM: Gossypium hirsutum

; FEATURE:
; NAME/KEY: unsure

; LOCATION: (1)..(564)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3825-027-Q6-K6-G2
US-10-021-323-3555

157 GAGGAGGGTATTAAACTGTGCAAGAGCGCGTAAAGTCAGGCATCAACTACATTCAGACT 216
166 TCCCCCTATTATGGAGGAACACTGTCTGAGAAATGCTTGGTAAAGGACATAAAGGCTTTG 225
217 GCTCCCTGGTATGGTCAGGGTCGCTCTGAGAGGTCCTGGACTGGCCCTAAAGATGTG 276
226 C-----AAGTCCTTAGAAGTGAATACATTTGGGCTACTAAGTGTGGTATAGATATAAGAA 279
277 CGCGGGGAATCCTACTATATCGCCACGAAAGTGGCTCGCTACGAACTGGACTACGATAA 336
280 GGTTTTGAATTCAGTCTCAGAGAGTAAAGAGATATTGACGAGAGCTTGGAGAGCTT 339
337 ATGTTTGAATTTAGTCTCCAGAGAGACGCGCGAAGCGTGGAGAGAGCTTGAACCTACTT 396
340 CAGCTTGATTTATGTTGCACATCTTCAATGCCATGACATTTAGTTGCG- --GGTCTCTTGAT 396
397 GGCTGGACTACGTTGATGTCTCCAGATTCAGATATCGAGTTGCGCAAGGATCTGGAC 456
397 CAGATTGTGAGTGAACAAATCTCTGCTCTTCAGAAACTGAACAGAGGGGAAGCCGG 456
457 ATTGTGATCAACGAGACACTGCCCACTTTGGAGCAGCTGCTCAAGGAGGGCAAGCAAGA 516
457 TTCAATGGTATCACTGGTCTTCGTTAGATATTTTCACTTATGTTTGCATCGAGTGCT 516
517 TTCAATGGAGTGTCCGCTTACCGGATTCGCGTCTTAAGGAGTTCCTGACCCGA---ACA 573
517 CAGGAGACTGTGATGTGATATTTGTCATCTGATTCAGGGGTTAATGATTCGAGCTTG 576
574 GCGGGAAGACTGATACGGTCTCACTATGCCAGATACACCTGACCGATGAACGCTC 633
577 CTGGATTACTACTTTACCTTTGAGAGCAAGTGTGGGTGTGATAGTCTTCTCATTA 636
634 CTGGAGTACTGATTTCTTCAAGTCCAGAACCTGGCGTCTATCTGTCCGCGAGTCTCAT 693
637 GCATGGGCTCTTTACAGACAAAGTCTCTCGATGGCAGCCCTGCTTCCCTGAGCTC 696
694 GCGCTCGAGCTGTGACCAATGCGGTCCAGCCATGCGTCCAGAGGAGTCCGGCCAGTATGAGCAG 753
697 AGTCTGCAAGCAAGCGGAGTGTCTCACTGCAATCAAGGGCGCAAGAGATCAAAAG 756
754 AAGGCCATTCGCGGAAGCATCGGAGTCTGCAAGGACGCGGTGGAGCTGGGCAAG 813
757 TTAGCTCTGCAATACA 772
814 CTGGCCATGTACTACA 829

RESULT 12
US-10-425-115-29198
; Sequence 29198, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 29198
; LENGTH: 342
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_126638C.1
US-10-425-115-29198

Query Match 10.4%; Score 99.4; DB 21; Length 342;
Best Local Similarity 59.3%; Pred. No. 3.1e-21;
Matches 169; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

282 TTTTGATTTAGTGTGCTGAGAGAGTAAGAAAGATATTACGAGAGCTTTGGAGAGCTTCA 341
48 TTTTGACTTTTCTCCAGCGGTGTGGCAGCATATTATAGGGAAAGCCTTTCCCGACTGG 107
342 GCTTGATTTATGTTGACATACCTTCAATGGCATGACATTTGAGTTCGGGTCTCTTGATCAGAT 401
108 GCTGAGTACCTCGCATGCTTTCAGCTCATGACATAGAGTTTCAACCCATCTCTTAACAGAT 167
402 TGTGAGTCAAAACAATTCCTGCTCTTCAGAAACTGAAACAAAGAGGGAAGACCCCGTTCAT 461
168 TGTGAATGAGACAATTTCTCTCCAAAAGTCCATGGGAACGGGTAGGACGGCTCAT 227
462 TGTATCACTGCTCTTCGGTTAGATATTTTCACTTATGTTTGTGATGAGTGCCTCCAGG 521
228 TGGCATTACCGGTTGGCTCTCAAAATCTACCTTTATGGGCTTGACCGGTAGAACCAAG 287
522 GACTGTGATGATGATTTGTCTATCTGTCATACGGCGTTAATGA 566
288 CTTGGTGGCGGATTTCTATTGTTGTTACTACGGGATCCAAAG 332

RESULT 13
US-11-097-143-40478
; Sequence 40478, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40478
; LENGTH: 1051
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-40478

Query Match 10.2%; Score 98.2; DB 26; Length 1051;
Best Local Similarity 49.9%; Pred. No. 1.6e-20;
Matches 359; Conservative 0; Mismatches 346; Indels 12; Gaps 4;
63 TGTGTCCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCCGCGGAAGATGATCGCGTCCGCAC 122
142 TGTGTCCTCTCTCCAAAGCTCTTCTCCGATGACTTCGATCCGAGGAGGCGATCCTCAC 201
123 COTGGCGAGGCTTTCCGCTCTCGGTATCACTTCTTCACACCTCCCGGTATATATGAGG 182
202 GGTGAGGAAGCCATTAGATCCGCTATCACTACATACACACGCGTCCCTTTTATGGCA 261
183 AACACTGTCTGAGAAATGCTTTGGTAAGGACTAAAGGCTTTGCG-----AAGTCCCTAGA 237

Db 262 AGGCAAAATCGAAGAGCTCTTGGCCAGCGGCTCAAGGATGTGCCCCGGAGGCGCTATTATA 321
 QY 238 AGTGACTACATTTGGCTACTAAGTGTGTAGATATAAAGAGGT-TTTGATTTTCAGTGC 296
 Db 322 TATAGCACTAAAGTTGACAGTTACGAGTTGGATCCAAACATATGTTCCGACTATACGGC 381
 QY 297 TGAGAGAGTAAGAAAGATATTGACGAGAGCTTTGAGAGGCTTCAGCTTGATATTGTTGA 356
 Db 382 TGCCAAAGGCTCGGGAGAGTGTGAAGCGTAGTCTGGAGCTGCTCCAGTTTGGACAGGGTGA 441
 QY 357 CATACTTCATTGCCATGACATTTGAGTTGGGTC---TCTTGATCAGATTGTGAGTGAAC 413
 Db 442 CGTACTACAGGTTTCATGAGCTGGATCGCGCACCTAGTCTGGACATGGTGTGCTGAATGAGAC 501
 QY 414 AATTCTGCTCTTCAGAACTGAAACAGAGGGGAGAGCCCGTTTCATTGGTATCACTGG 473
 Db 502 CATACCCGCTCGAGGAGTACGTCCAGCGGGAAGGCTCGATTTCATCGGAGTCAACGC 561
 QY 474 TCTTCGTTAGATATTTTCACTTATGTTTCTTGAATGAGTGCCTCAGGAGCTGTGATGT 533
 Db 562 CTACGATGTGAGCGTGTGAAGGAGTGTGCCGAGCG--GGGCAAAAGTGTGCATCCAGGT 618
 QY 534 GATATTGTCATCTGCTTATTCAGGCTTAAATGATTTCGAGCTTGTGCTGATTTACTACTTA 593
 Db 619 GGTGCTCAACTATGCCCGCTACACCCCTGTGAGCAACACTTTGCTGCGCCACATGAAGGC 678
 QY 594 CTTGAGAGCAAGGTGTGGGTGTGATAGTCTTCTCCATPAGCAATGGGCTCTCTTAC 653
 Db 679 CTTCCAGGAGATGGCGTGGGCTGTCTGTGCGCGGCGCACTCTTGGGACTCTTAAG 738
 QY 654 AGAACAAGTCTCTCTGAATAGCAACCTCTTCCCTGAGCTCAAGTCTGCAAGCAAGC 713
 Db 739 CAACGCTGAGCCCACTGCTGGCATCTGCTAGTCCGGAACCTCTAGCTGTGGGCAAGC 798
 QY 714 CCGAGTGTCTACTGCAATCAAGGCAAGGATCACAAGATTAGCTCTGCAATACA 772
 Db 799 GGGAGCCGAATCTGTGCAAGAGAACTTGAGCTTGGAAAGCTGGCCATGTACTATA 857

RESULT 14

US-10-021-323-2292
 ; Sequence 2292, Application US/10021323
 ; Publication No. US20040123340A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Deikman, Jill
 ; APPLICANT: Feng, Paul C.C.
 ; APPLICANT: Pincher, Karen L.
 ; APPLICANT: Ziegler, Todd E.
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE REFERENCE: 38-21(52274)B
 ; CURRENT APPLICATION NUMBER: US/10/021,323
 ; PRIOR APPLICATION NUMBER: 2001-12-12
 ; PRIOR FILING DATE: 2000-12-14
 ; NUMBER OF SEQ ID NOS: 17880
 ; SEQ ID NO 2292
 ; LENGTH: 570
 ; TYPE: DNA
 ; ORGANISM: Gossypium hirsutum
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (1)...(570)
 ; OTHER INFORMATION: unsure at all n locations
 ; OTHER INFORMATION: Clone ID: LIB3825-012-Q1-K6-C6
 US-10-021-323-2292

Query Match 10.0%; Score 95.8; DB 20; Length 570;
 Best Local Similarity 72.5%; Pred. No. 6.5e-20;
 Matches 124; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
 QY 5 CGAAATAGAGCTTTCAGCTTTGGGGAACACAGGGCTTAAGGTTAGCGCGTGTGTTTG 64

Db 261 CCAAACTAGAAATGAGACCTCTCGAAACACGGGCTTAAAGCTCAGTAGCGTGGCTTCG 320
 QY 65 GTGCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGAAGATGATGCCGTCGCCACCG 124
 Db 321 GAGCTTCTCTCTCGGAGCGTCTTCGGTTCGGTTTCGAAAGCGAGCTGTGCGCTCG 380
 QY 125 TGCSCGAGGCTTTCGGTCTCGGTATCAACTTCTTCGACACCTCTCCCGTATT 175
 Db 381 TCCTCGAAACCTTCGCGCTCTGAATCAACTTCTTCGACACCTCTTCGTAAT 431

RESULT 15

US-10-425-114-23348
 ; Sequence 23348, Application US/10425114
 ; Publication No. US20040034888A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Jingdong
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Screen, Steven E.
 ; APPLICANT: Tabaska, Jack E.
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE REFERENCE: 38-21(53313)B
 ; CURRENT APPLICATION NUMBER: US/10/425,114
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 73128
 ; SEQ ID NO 23348
 ; LENGTH: 255
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: LIB3595-041-P7_FLI
 US-10-425-114-23348

Query Match 9.8%; Score 94; DB 19; Length 255;
 Best Local Similarity 66.0%; Pred. No. 1.5e-19;
 Matches 136; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
 QY 10 ATAGAGCTTCAGCTTTGGGGAACACAGGGCTTAAGGTTAGCGCGTGTGTTTGGTGCC 69
 Db 7 ATGAGCTCCCGAGCTGGCGGTACGGGTCTCCGCGTCAGCGCGTTCGGCTTTGGTGCC 66
 QY 70 TCTCGCTCGGAAGTGTCTTCGGTCCAGTCCCGAAGATGATGCGCTCGCCACCGTGGCC 129
 Db 67 TCTCGCTTGGCAACGTTTTCGGGACGTGTCGCCCGGACACCGCGCGCGCTCGCC 126
 QY 130 GAGCTTTCGCTCGGTATCAACTTCTTCGACACCTCTCCCGTATTATGAGGGAACATG 189
 Db 127 CGCGGCTCGACCTCGGATCAACTTCTTCGATACCTCAGCGTACTACCGTGGCAGATA 186
 QY 190 TCTGAGAAATGCTTGTGTAAGGACT 215
 Db 187 TCGAGTCAGTCTCGCGGATTGCT 212

Search completed: November 13, 2005, 08:23:10
 Job time : 957 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2005, 05:27:45 ; Search time 3734 Seconds
(without alignments)
9786.207 Million cell updates/sec

Title: US-10-606-300-12

Perfect score: 960

Sequence: 1 atgacgaaatagacttcg.....gtggaatccatcagaactaa 960

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gss1:*

9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	820.2	85.4	894	5	BX839165
2	625.4	65.1	638	1	AV825911
3	554.6	57.8	953	7	CK253464
4	536	55.8	998	7	CK251403
5	533.8	55.6	987	7	CK250147
6	527.2	54.9	979	7	CK282536
7	520	54.2	948	7	CK243867
8	510.2	53.1	936	7	CK257246
9	504.8	52.6	945	7	CK245431
10	497.8	51.9	883	7	CK286730
11	490.6	51.8	922	7	CK284486
12	479	49.9	926	7	CK282535
13	468.6	48.8	903	7	CK282536
14	463.2	48.2	859	7	CK297635
15	459.4	47.9	828	6	CD576576
16	452.6	47.1	912	7	CK243866
17	451.6	47.0	926	7	CK250148
18	445.4	46.4	822	7	CK366839
19	440.6	45.9	741	7	CF568880
20	434	45.2	826	7	CK297748
21	433.4	45.1	814	7	CK294400
22	429.2	44.7	1302	9	CL379301
23	428	44.6	858	7	CK282537
24					

25	427.6	44.5	1287	3	AY106400
26	425.2	44.3	817	7	CK284947
27	423	44.1	695	7	CV430394
28	422	44.0	777	7	CK254163
29	420.4	43.8	698	4	BI920114
30	419.6	43.7	670	5	B0888949
31	418	43.5	719	7	CF417257
32	416.8	43.4	834	7	CO363121
33	414.4	43.2	836	7	CK297749
34	413.4	43.1	661	5	BX835470
35	404.2	42.1	655	7	CK758916
36	402.4	41.9	834	7	CK286442
37	394.8	41.1	829	7	CK291957
38	393.6	41.0	739	4	BI931318
39	390	40.6	706	4	BG646660
40	385	40.1	806	7	CO099522
41	384.8	40.1	702	4	BI416734
42	382	39.8	739	5	BQ797990
43	369.2	38.5	810	6	CB618586
44	366.6	38.2	780	7	CK251404
45	365	38.0	709	5	BQ986492

ALIGNMENTS

RESULT 1
BX839165
LOCUS BX839165 894 bp mRNA linear EST 11-FEB-2004
DEFINITION BX839165 Arabidopsis thaliana Adult vegetative tissue Col-0
Arabidopsis thaliana cDNA clone GSLTSL582B09 SPRIM, mRNA sequence.
ACCESSION BX839165
VERSION BX839165.1 GI:42533248
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 894)
AUTHORS Castell, V., Aury, J.M., Jaillon, O., Wincker, P., Clepet, C., Menard, M., Cruaud, C., Quetier, F., Scarpeilli, C., Schachter, V., Temple, G., Caboche, M., Weissenbach, J., and Salanoubat, M.
TITLE Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences: A Combined Approach to Evaluate and Improve Arabidopsis Genome Annotation
JOURNAL Unpublished (2004)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
The sequences are based on single pass reads.
Life Technologies (a division of Invitrogen) members carried out full-length libraries construction : Temple G.
Genoscope members carried out sequencing and annotation : Castell V., Aury J.M., Jaillon O., Wincker P., Menard M., Cruaud C., Schachter V., Weissenbach J., Salanoubat M.
URV INRA : Clepet C., Caboche M.
Annotation is based on the June 2003 version of the Arabidopsis genome released by MIPS (Munich Information center for Protein Sequences).
http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EP/EST
http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.

FEATURES
source
Location/Qualifiers
1..894
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/ecotype="Col-0"
/db_xref="taxon:3702"
/clone="GSLTSL582B09"
/tissue_type="Adult vegetative tissue"
/clone_lib="Arabidopsis thaliana Adult vegetative tissue Col-0"

ORIGIN									
Query Match 85.4%; Score 820.2; DB 5; Length 894; Best Local Similarity 97.8%; Pred. No. 2.9e-247; Matches 842; Conservative 0; Mismatches 18; Indels 1; Gaps 1;									
QY	1	ATGACGAAATAGAGCTTCGAGCTTTGGGGAACACAGGCTTAAGTTAGCCCGCTTGGT	60						
Db	34	ATGACGAAATAGAGCTTCGAGCTTTGGGGAACACAGGCTTAAGTTAGCCCGCTTGGT	93						
QY	61	TTTGGTGCCTCCGCTCGGAAGTCTTCGGTCCAGTCGCCGAA-GATGATGCCGTGCG	119						
Db	94	TTTGGTGCCTCCGCTCGGAAGTCTTCGGTCCAGTCGCCGAAAGATGATGCCGTGCG	153						
QY	120	CACCGTGCAGGCTTTCCGCTCGGTATCAACTTCTTCGACACCTCCCGCTATTATGG	179						
Db	154	CACCGTGCAGGCTTTCCGCTCGGTATCAACTTCTTCGACACCTCCCGCTATTATGG	213						
QY	180	AGGAACACTGTCTGAGAAATGCTTGGTAAAGGACTAAAGGCTTTGCAAGTCCCTAGAG	239						
Db	214	AGGAACACTGTCTGAGAAATGCTTGGTAAAGGACTAAAGGCTTTGCAAGTCCCTAGAG	273						
QY	240	TGACTACATTTGGCTACTAAGTGTGTAGATATAAGAGGTTTGTATTCAGTGTGA	299						
Db	274	TGACTACATTTGGCTACTAAGTGTGTAGATATAAGAGGTTTGTATTCAGTGTGA	333						
QY	300	GAGAGTAAGAAAGAGTATTGACGAGAGCTTGGAGAGGCTTCAGCTTGATTATTTGACAT	359						
Db	334	GAGAGTAAGAGAGTATTGACGAGAGCTTGGAGAGGCTTCAGCTTGATTATTTGACAT	393						
QY	360	ACTTCATTCGCATGACATTTAGTTCGGGTCTCTTGATCAGATGTGAGTGAACAATTC	419						
Db	394	ACTTCATTCGCATGACATTTAGTTCGGGTCTCTTGATCAGATGTGAGTGAACAATTC	453						
QY	420	TGCTTTTCAGAACTCAACACAGAGGGGAGACCCGGTTTCATTTGATCAGTGTCTCC	479						
Db	454	TGCTTTTCAGAACTCAACACAGAGGGGAGACCCGGTTTCATTTGATCAGTGTCTCC	513						
QY	480	GTTAGATATTTTACATTTATTTTTCGATCGAGTGCCTCCAGGAGCTGTGATGTGATTT	539						
Db	514	GTTAGATATTTTAACTTATTTATTTGATCGAGTGCCTCCAGGAGCTGTGATGTGATTT	573						
QY	540	GTCATCTGATTCACGGGTTAATGATTCGAGGTTGCTGGATTTACTACCTTACTGAA	599						
Db	574	GTCATCTGATTCACGGGTTGATGATTCGAGGTTGCTGGATTTACTACCTTACTGAA	633						
QY	600	GAGCAAGGTGTGGGTGTGATAGTCTTCTCATTAGCAATGGGCTCTCTTACAGAA	659						
Db	634	GAGCAAGGTGTGGGTGTGATAGTCTTCTCATTAGCAATGGGCTCTCTTACAGAA	693						
QY	660	AGGTCTCTCTGATGACCCCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAAAGCCGAGT	719						
Db	694	AGGTCTCTCTGATGACCCCTGCTTCCCTGAGCTCAAGTCTGCGAGCGAAGCCGAGT	753						
QY	720	TGCTCTCTGCAATCAAAAGGCGAAGATCAAAAGTTAGCTTGCATATACAGTTAGC	779						
Db	754	TGCTCTCTGCAATGAGAGGCGAAGATCAAAAGTTAGCTTGCATATACAGTTAGC	813						
QY	780	AAACAAGGAGATTTCCGTCGGTGTGTGGATGAGCTCTGTCTCAGGTAGAGAAA	839						
Db	814	AAACAAGGAGATTTCCGTCGGTGTGTGGATGAGATATGCTCAGGTAGAGAAA	873						
QY	840	TGTTGCAGCAGTTTACAGAGCT	860						
Db	874	TGTTGCAGCAGTTTACAGAGCT	894						
RESULT 2									
AV825911 638 bp mRNA linear EST 01-APR-2002									
LOCUS AV825911 RAFL7 Arabidopsis thaliana cDNA clone RAFL07-17-C15 5',									
DEFINITION mRNA sequence.									
ACCESSION AV825911									

AV825911.1	GI:19867971
EST.	
Arabidopsis thaliana (thale cress)	
Arabidopsis thaliana	
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;	
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.	
1 (bases 1 to 638)	
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,	
Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,	
Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.	
and Shinozaki, K.	
Large scale analysis of Arabidopsis full-length cDNA (2002b)	
Unpublished (2002)	
Contact: Motoaki Seki	
Plant Functional Genomics Research Group	
RIKEN Genomic Sciences Center	
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan	
Tel: 81-298-36-4359	
Fax: 81-298-36-9060	
Email: msekiorc.riken.go.jp	
An Arabidopsis full-length cDNA library was constructed essentially	
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI	
and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et	
al., submitted for publication) digested with BamHI and SalI. This	
clone is in a modified pBluescript vector. Please visit our web	
site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further	
details.	
Location/Qualifiers	
1..638	
/organism="Arabidopsis thaliana"	
/mol_type="mRNA"	
/db_xref="taxon:3702"	
/clone="RAFL07-17-C15"	
/dev_stage="rosette plants"	
/lab_hosts="DH10B"	
/clone_lib="RAFL7"	
/note="Site 1: BamHI; Site 2: SalI; subjected to	
cold-treated (1, 2, 5, 10, 24 hr)"	

Query Match	65.1%;	Score 625.4;	DB 1;	Length 638;
Best Local Similarity	99.4%;	Pred. No. 9.9e-186;		
Matches 626;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;
QY	1	ATGACGAAATAGAGCTTCGAGCTTTGGGGAACACAGGCTTAAGTTAGCCCGCTTGGT	60	
Db	9	ATGACGAAATAGAGCTTCGAGCTTTGGGGAACACAGGCTTAAGTTAGCCCGCTTGGT	68	
QY	61	TTTGGTGCCTTCGCTCGGAAGTGTCTTCGGTCCAGTCGCCGAAAGATGATGCCGTGCG	120	
Db	69	TTTGGTGCCTTCGCTCGGAAGTGTCTTCGGTCCAGTCGCCGAAAGATGATGCCGTGCG	128	
QY	121	ACCGTGCAGGCTTTCCGCTCGGTATCAACTTCTTCGACACCTCCCGCTTATATGGA	180	
Db	129	ACCGTGCAGGCTTTCCGCTCGGTATCAACTTCTTCGACACCTCCCGCTTATATGGA	188	
QY	181	GGAAACACTGTCTGAGAAAATGCTTGGTAAAGGACTTAAAGGCTTTTCAAGTCCCTAGAAGT	240	
Db	189	GGAAACACTGTCTGAGAAAATGCTTGGTAAAGGACTTAAAGGCTTTTCAAGTCCCTAGAAGT	248	
QY	241	GACTACATTTGGCTTACTAAGTGTGTAGATATAAAGAGGTTTTTGAATTCAGTCTGAG	300	
Db	249	GACTACATTTGGCTTACTAAGTGTGTAGATATAAAGAGGTTTTTGAATTCAGTCTGAG	308	
QY	301	AGAGTAAGAAAGAGTATTGACGAGAGCTTGGAGAGGCTTCAGCTTGATTATTTGACATA	360	
Db	309	AGAGTAAGAAAGAGTATTGACGAGAGCTTGGAGAGGCTTCAGCTTGATTATTTGACATA	368	
QY	361	CTTCAATGCCATGACATTTAGTTCGGGTCTCTTGATCAGATTTGTGAGTGAACAATTCCT	420	
Db	369	CTTCAATGCCATGACATTTAGTTCGGGTCTCTTGATCAGATTTGTGAGTGAACAATTCCT	428	

QY 421 GCTCTTCAGAACTGAAACAAGAGGGGAAGACCCGGTTCATTGGTATCAGTGGTCTCCG 480
 Db 429 GCTCTTCAGAACTGAAACAAGAGGGGAAGACCCGGTTCATTGGTATCAGTGGTCTCCG 488
 QY 481 TTAGATATTTTCACTATGTTCTTGATCGAGTGCCTCCAGGACTGTCGATGATATG 540
 Db 489 TTAGATATTTTCACTATGTTCTTGATCGAGTGCCTCCAGGACTGTCGATGATATG 548
 QY 541 TCATCTGTCATTACGGGTTAATGATTCGAGCTTGTCTGGATTTTACTACCTTACTTGAAG 600
 Db 549 TCATCTGTCATTACGGGTTAATGATTCGAGCTTGTCTGGATTTTACTACCTTACTTGAAG 608
 QY 601 AGCAAAAGTGGGTGATGATAGTGCCTTCT 630
 Db 609 AGCAAAAGTGGGTGATGATAGTGCCTTCT 638

RESULT 3 CK253464

LOCUS EST737101 potato callus cdna library, normalized and full-length
 DEFINITION Solanum tuberosum cdna clone POC562 5' end, mRNA sequence.

ACCESSION CK253464

VERSION CK253464.1 GI:39808422

KEYWORDS EST.

SOURCE Solanum tuberosum (potato)

ORGANISM Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamids; Solanales; Solanaceae; Solanum.

REFERENCE 1 (bases 1 to 953)

BUELL, C.R., HART, A., ZISMANN, V., KARAMYCHEVA, S.A. and BAKER, B.

Generation of ESTs from potato callus tissue

Unpublished (2003)

Contact: Robin Buell

The Institute for Genomic Research

9712 Medical Center Dr, Rockville, MD 20850, USA

Email: potato-array@tigr.org

Clones can be requested from the University of Arizona Genomics

Institute via <http://genome.arizona.edu/orders/>

Seq primer: ATT TAG GTG ACA CTA TAG.

FEATURES source

Location/Qualifiers

1..953

/organism="Solanum tuberosum"

/mol_type="mRNA"

/cultivar="Kennebec"

/db_xref="taxon:4113"

/clone="POCC562"

/tissue_type="callus"

/lab_host="DH10B-Tona"

/clone_lib="potato callus cdna library, normalized and

full-length"

/note="vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;

supplier: RNA was isolated from Solanum tuberosum var.

Kennebec callus tissue grown on solid media."

ORIGIN

Query Match 57.8%; Score 554.6; DB 7; Length 953;
 Best Local Similarity 74.2%; Pred. No. 2.8e-163;
 Matches 701; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

QY 6 GAAATAGAGCTTCGAGCTTTGGGGAACACAGGGGCTTAAGGTAGCGCGTGTGGTTGG 65
 Db 6 GACATTGAGCTCCGACCACTGGGCAACATGGACTCAATCTAGCTCTGCGGTTTCGG 65
 QY 66 TGCCTCTCCGTCGGAAGTGTCTTCGGTCCAGTCCGCAAGATGATCCGTCGCCACCGT 125
 Db 66 CGCTTCTCTCCGCAAGGTTTTCGGGATGTCTCCGCAACAGAGCGCTTCGCGCGGT 125
 QY 126 GCGGAGCGCTTCGCTCTCGGTATCAACTCTTCGACACCTCCCGTATTATGAGGAGAC 185
 Db 126 CGGGAAGCGCTTCGCTCTCGGTATCAACTCTTCGACACCTTCGCGTATTATGAGGAGAC 185

QY 186 ACTGTCGAGAAAATGCTTGGTAAGGACTAAAGGCTTTTGAAGTCCCTAGAGTACTA 245
 Db 186 GTTATCGGAAAAGGTACTAGGAAAAGCTTTTGAAGGCTCTTTGGAGTACTCAGATCAGTA 245
 QY 246 CATTGTGGCTTACTAAGTGTGTAGATATAAAGAAAGGTTTTTGAATTTCAAGTGTCTGAGAGAGT 305
 Db 246 CATTGTGTCTACAAAATGTGGAGGTACAAAGAGGGAATTTGAATTTTGTAGTGTCTGAGAGAGT 305
 QY 306 AAGAAAAGATATTGACGAGAGCTTTGGAGAGGCTTACGCTTGAATTTATGATGATGATTTCA 365
 Db 306 GACTAAAAGTATTGATGAGAGCTTTGGAGAGGCTGCAGCTTGAATTTATGATGATGATTTCA 365
 QY 366 TTGCCATGACATTGAGTTCGGGTCTCTTGATCAGATGTTGAGTGAACAATTTCTGCTCT 425
 Db 366 ATGTCATGATATTGAATTTGGGTCCCTCGATCAGATTTGAAATGAGACACTTCCCGCCT 425
 QY 426 TCAGAAACTGAAACAAGAGGGGAAGACCCGGTTCATTGGTATCACTGGTCTTCGGTTAGA 485
 Db 426 TCTAAAATTTGAGCAAGCTGGAAGATCCGCTTCATTGGTATTAACCGGCTTACCTTTGGG 485
 QY 486 TATTTTCACTTATGTTCTTGATCGAGTGCCTCCAGGAGCTGTTCGATGTGATTTGTGCATA 545
 Db 486 GATATTCACTTATGTTCTTGATCGGTCCCTCCAGGCAAGTTGACGTTATCTCTGTGCATA 545
 QY 546 CTGTCATTACGGCTTAAATGATTCGACGTTGCTGGATTACTACTTACTTGAAGAGCAA 605
 Db 546 TTGTCACACTACAGTGTCAACGATTTCACTTTGGAGGATCTGTTGCCATACCTTGAAGAGCAA 605
 QY 606 AGGTGTGGGTGATAGTATGTCCTTCCATTAGCAATGGGCTCTTACAGAACCAAGGTCC 665
 Db 606 GGGTGTGGGGGTGATCAGTGTCTCTCTTCAATGGGTCTTCTTACTGAGGCTGTGG 665
 QY 666 TCCTGAATGGCACCCCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAAAGCCGAGTTGCTCA 725
 Db 666 TCCAGAGTGGCACCCCTGCTTCTTCTGAACCTTAAGGCTGCTGCGGAGCTGCTGCTGATCA 725
 QY 726 CTGCAAAATCAAAAGGGCAAGAGATCAAAAATTTAGTCTCTGCAATACAGTTAGCAAAACA 785
 Db 726 CTGCAAAAGGAAGGAAAGAAATATCTCAAAATTTAGCTTGCAGTACAGCTTTAGCAAAATAC 785
 QY 786 GGAGATTTCTCGGTGTGGTGGATGAGCTCTGCTCACAGGTAGCAAGAAATGTTGC 845
 Db 786 CGATATTTCTTCCATCTAGTGGGAATGAAGTCTGTTAAAGAGGTGGAGGAAATATAGC 845
 QY 846 AGCAGTTTACAGAGCTTCAAAAGTCTGGGGAGTGAAGCAAAACTCTGTCTGAGGTTGAAGC 905
 Db 846 AGCTGCCCTAGACTAGAAACGCTGGAATGGATGAAGAAACATTATCAGAGATCACAGA 905
 QY 906 TATTCTCGAGCCTGTAAAGAAATCTGACATGCCCAAGTGGAAATCCA 950
 Db 906 CATTCTGAACCAAGTGAAGAAATCAGACATGGGCTAGCGGTATCAAA 950

RESULT 4

CK251403

LOCUS

DEFINITION

EST735040 potato callus cdna library, normalized and full-length

Solanum tuberosum cdna clone POCBN18 5' end, mRNA sequence.

CK251403

VERSION CK251403.1 GI:39804397

KEYWORDS EST.

SOURCE Solanum tuberosum (potato)

ORGANISM Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

asterids; lamids; Solanales; Solanaceae; Solanum.

REFERENCE 1 (bases 1 to 998)

BUELL, C.R., HART, A., ZISMANN, V., KARAMYCHEVA, S.A. and BAKER, B.

Generation of ESTs from potato callus tissue

Unpublished (2003)

Other ESTs: EST735041

Contact: Robin Buell

The Institute for Genomic Research

9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
Clones can be requested from the University of Arizona Genomics
Institute via <http://genome.arizona.edu/orders/>.
Seq primer: ATT TAG GTG ACA CTA TAG.

FEATURES

Location/Qualifiers
1..998
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="POCEN18"
/tissue_type="callus"
/lab_host="DH10B-Tona"
/clone_lib="potato callus cDNA library, normalized and
full-length"
/note="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
supplier: RNA was isolated from Solanum tuberosum var.
Kennebec callus tissue grown on solid media."

ORIGIN

Query Match 55.8%; Score 536; DB 7; Length 998;
Best Local Similarity 74.2%; Pred. No. 2.2e-157;
Matches 677; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

QY 6 GAAATAGAGCTTCGAGCTTTGGGAAACACAGGGCTTAAGTTAGCGCGTGTGGTTTGG 65
DB 81 GACATTGAGCTCCGATCACTGGGCAACACTGACTCAATCTCAGCTCTGTGGGTTGG 140
QY 66 TGCCTCTCGCTCGAAGTGCTTCGCTCCAGTCGCGCAAGATGATCGCTGCCACCGT 125
DB 141 CGCTTCTCTCTCGGCAAGTTTTCGGCGATGCTCCGAAACAGAGCGCTTCGCGCGCT 200
QY 126 GCGGAGGCTTTCGCTCGGTATCAACTTCTTCGACACCTCCCGTATTATGAGGAAAC 185
DB 201 CGGGAAGCTTTCGCTCGGCTCAATTTCTTTCGACACTTCCCGTATTATGAGGAAAC 260
QY 186 ACTGTCTGAGAAATGCTTGTGTAAGGACTAAAGGCTTTGCAAGTTCCTAGAGTACTA 245
DB 261 GTTATCGGAAAGGTAAGTACTAGGAAAGCTTTGAAGGCTCTTGAGTACCTAGAGTAC 320
QY 246 CATTGTGCTACTAAGTGTGGTAGATATAAGAAAGTTTTCAGTTTCACTGCTGAGAGT 305
DB 321 CATTGTGCTACAAATTTGGGAGGTACAAAGAGGATTTGATTTAGTGTGAGAGT 380
QY 306 AAGAAGAGTATTGACGAGCTTCGAGAGGCTTCAGCTTGAATTTGATGATGATTTCA 365
DB 381 GACTAAAGTATTGATGAGCTTCGAGAGGCTGAGCTTGAATTTGATGATGATTTCA 440
QY 366 TTGCCATGACATGATGCTCGGCTCTTGTGATCAGATTTGTGAGTGAACAAATTCCTGCT 425
DB 441 ATGTCATGATATTGAATTTGGGTCCTCGATCAGATTTGTAATGAGACACTTCCGCGCT 500
QY 426 TCAGAACTGAACAGAGGGAAGACCGGTTCACTGATCACTGCTTCGTTAGA 485
DB 501 TCTAAATTTGAAGCAAGCTGGAAGATCCGCTTCAATTTGATTAACCGGCTTACCTTTGG 560
QY 486 TATTTTCACTATTGTTTGTGATGCTCCAGGAGCTTCGAGTGTGATTTGATTTGATTA 545
DB 561 GATATTCACTATTGCTTGTGATCGGCTCCCTCAGGACAGTTGATGATCATCTGCTATA 620
QY 546 GTGTCAATTACGGCGTTAATGATTCAGCGTTGTGATTTTACCTTACTTTGAAGAGCAA 605
DB 621 TTGTCACTACTGTCAACGATTTCAACTTTGGAGGATCTGTTCATCACTACCTGAAGAG 680
QY 606 AGGTGTGGTGTGATGAAGTCTTCTTCAATAGCAATGGGCTTCCTTACAGAACAGGTCC 665
DB 681 GGGTGTGGGGGTGATCAGTGCTTCTCTCTTTCAATGGGTCTTCTTACTAGGCTGGTGG 740
QY 666 TCCCTGAATGGCACTTCCCTGCTCAAGTCAAGTCTGCAAGCAAGCGGCTGCTGCTCA 725
DB 741 TCCAGATGGCACTTCTTCTGMACTTAAGGCTGCTTCCGAGGCTGCTGCTGATCA 800

QY 726 CTGCAATCAAGGGCAAGAGATCAAAAGTTAGCTCTGCAATACAGTTTAGCAACAA 785
DB 801 CTGCAAGGAAGGAAGAAATATCTCAAAATTAGCTTGCAGTACAGCTTAGCAATAC 860
QY 786 GGAGATTTCGTCGGTGTGGTGGATGAGCTCTGCTCACAGGTAGAGAAATTTGTC 845
DB 861 CGATATTTCTTCCATCTAGTGGGATGAAGTCTGTTAAGAGGTGGAGGAATATAGC 920
QY 846 AGCAGTTACAGAGCTTGAAGTCTGGGATGGATCAAGAACTCTGTCTGAGGTGAAGC 905
DB 921 AGCTGCCCTAGACTAGCAACGGCTGGAATGGATGAAGAAACATTATCAGAGATCACAGA 980
QY 906 TATTCTGAGCC 917
DB 981 CATTCTGGAAC 992

RESULT 5

CK250147
LOCUS
DEFINITION
EST733784 potato callus cDNA library, normalized and full-length
Solanum tuberosum cDNA clone POCBD80 5' end, mRNA sequence.

CK250147

CK250147.1 GI:39801934

EST.

Solanum tuberosum (potato)

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum.

1 (bases 1 to 987)

Buell, C.R., Har, A., Zismann, V., Karamycheva, S.A. and Baker, B.

Generation of ESTs from potato callus tissue

Unpublished (2003)

Other ESTs: EST733785

Contact: Robin Buell

The Institute for Genomic Research

9712 Medical Center Dr, Rockville, MD 20850, USA

Email: potato-array@tigr.org

Clones can be requested from the University of Arizona Genomics

Institute via <http://genome.arizona.edu/orders/>.

Seq primer: ATT TAG GTG ACA CTA TAG.

FEATURES

Location/Qualifiers

1..987

/organism="Solanum tuberosum"

/mol_type="mRNA"

/cultivar="Kennebec"

/db_xref="taxon:4113"

/clone="POCBD80"

/tissue_type="callus"

/lab_host="DH10B-Tona"

/clone_lib="potato callus cDNA library, normalized and

full-length"

/note="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;

supplier: RNA was isolated from Solanum tuberosum var.

Kennebec callus tissue grown on solid media."

ORIGIN

Query Match 55.6%; Score 533.8; DB 7; Length 987;
Best Local Similarity 74.4%; Pred. No. 1.1e-156;
Matches 673; Conservative 0; Mismatches 232; Indels 0; Gaps 0;

QY 6 GAAATAGAGCTTCGAGCTTTGGGAAACACAGGGCTTAAGTTAGCGCGTGTGGTTTGG 65
DB 83 GACATTGAGCTTCGATCACTGGGCAACACTGGAAGTCTCAATCTCAGCTCTGTGGGTTGG 142
QY 66 TGCCTCTCGCTCGAAGTGTCTTCGCTCCAGTCCGGAAGATGATGCGTCGCCACCGT 125
DB 143 CGCTTCTCTCTCGGCAAGGTTTTCGGCGATGTCTCGAAACAGAGCTTTCGCCGCGT 202
QY 126 GCGGAGGCTTTCGCTCTCGGTATCAACTTCTTCGACACCTCCCGTATTATGAGGAAAC 185
DB 203 CCGGAGGCTTTCGCTTCGGCTCAATTTCTTTGACACTTCCCGTATTATGAGGAAAC 262

QY	186	ACTGTCTCAGAAAAATGCTTGGTTAAGGACTTAAGGCTTTGCCAAGTCCCTTAGAAGTGCATA	245
Db	263	GTATTTCGAAAAAGGTACTAGGCGAAAGCTTTGAAGGCTCTTTGGAGTAGTACCTAGAGATCAGTA	322
QY	246	CATTCTGGCTACTAAGTGTGTGATATAAAGAAAGGTTTTGATTTTCAGTGTCTGAGAGAGT	305
Db	323	CATTGTGTCTACAAAATTTGGGAGGTCAAGAGGGGATTTGATTTTAGTGTCTGAGAGAGT	382
QY	306	AAGAAAGAGTATTGCACGAGCTTTGGAGAGCTTCAGCTTGAATATATGTTGACATACTTCA	365
Db	383	GACTAAAGTATTGATGAGACTTTGGAGAGGCTGCAGCTTGATATGTTGATATGTTACA	442
QY	366	TTGCCATGACATTTGATTTGGGTCTCTTGATTCAGATTTGAGTGAACAAATTCCTGCTCT	425
Db	443	ATGTCATGATATTGAATTTGGGTCCCTCGATCAGATTTGAAATGAGACACTTCCGGCCCT	502
QY	426	TCAGAAACTGAAAACAAGAGGGGAGACCCGGTTCATTTGGTATCACCTGGTCTTCCCGTTAGA	485
Db	503	TCTAAAATTGAAGCAAGCTGGAAGATCCGCTTCATTTGGTATTAACCGGCTTACCTTTGGG	562
QY	486	TATTTTTCACCTATGTTCTTTGATCGAGTGCCTCCAGGGACTGTCGATGTGATATTGTTCATA	545
Db	563	GATATTTCACCTTATGTGCTTGATCGGCTCCCTCCAGGCCACAGTTGATGTCATCTCTGTTCATA	622
QY	546	CTGTCAATTACGGCGTTAATGATTCGAGCTTGCTGGATTTACTACTTACTTGAAGAGCAA	605
Db	623	TTGTCACTACTGTGTCAAGATTCAACTTTGGAGGATCTGTTGCCATACTTGAAGAGCAA	682
QY	606	AGGTGTGGGTGTGATTAAGTGTCTTCCATTAGCAATGGGCCCTCTTTACAGAAACAAGGTCC	665
Db	683	GGGTGTGGGGTGATCAGTGTCTTCTCTTTCAAATGGGTCTTCTTACTTGAGCCTGGTGG	742
QY	666	TCCTGAATGGCACCCCTGCTTCCCTTGAGCTCAAGTCTGCAAGCAAAAGCCGAGTGTGCTCA	725
Db	743	TCCAGAGTGGCACCCCTGCTTCTTCTGAACCTTAAGGCTGCCTGCCGAGCTGCTGCTGATCA	802
QY	726	CTGCAATCAAGGGCAGAGAGATCACAAAGTTAGCTCTGCAATACAGTTTACGAAACA	785
Db	803	CTGCAAGGAAAGGGAAGAATATCTCAAAATTAGCCCTTGCAGTACAGCTTAGCAAAATAC	862
QY	786	GGAGATTTCTGTCGGTGTGGTGTGGGATGAGCTCTGTCTCACAGGTAGAAAGAAAATGTTGC	845
Db	863	CGATATTTCTTCCATACTAGTGTGGGATGAGTCTGTTAAAGAGGTGGAGGAAAATATAGC	922
QY	846	AGCAGTTACAGAGCTTGAAAGTCTGGGGATGGATCAAGAAATCTCTGTCTGAGTGTGAAGC	905
Db	923	ASCTGCCCTAGAACTAGCAACGGCTGGGATGGATGAAGAAACATTATCAGAGATCACAGA	982
QY	906	TATTC	910
Db	983	CATTC	987

RESULT 6	CK282536	979 bp	mrna	linear	EST 02-AUG-2004
LOCUS	CK282536				
DEFINITION	EST745258 Nicotiana benthamiana mixed tissue cDNA library, normalized, full-length Nicotiana benthamiana cDNA clone NEWAD20 5', end, mRNA sequence.				
ACCESSION	CK282536				
VERSION	CK282536.1	GI:39854240			
KEYWORDS	EST.				
SOURCE	Nicotiana benthamiana				
ORGANISM	Nicotiana benthamiana				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamiids; Solanales; Solanaceae; Nicotiana.				
REFERENCE	1 (bases 1 to 979)				
AUTHORS	Buell, C.R., Hart, A., Zismann, V., Karamycheva, S.A., Day, B., Staskawicz, B., Jin, H. and Baker, B.				
TITLE	Generation of EST sequences from Nicotiana benthamiana				
JOURNAL	Unpublished (2003)				

Db 670 GGGTGGGAGTGAATCAGTGTCTCCCTCTTTGCAATGGGTCTTCTTACTAGGCTGGAGC 729
 Qy 666 TCCTGAATGGCACCCTGCTTCCCTCTGAGCTCAAGTCTGCAAGCAAGCCGAGTTGCTCA 725
 Db 730 TCCAGATGGCACCCTGCTTCTTGAACCTTAAGCTGCTTGGCGAGCTGCGTTGATCA 789
 Qy 726 CTGCAATCAAGGCGAAGAGATACAAAGTTAGTCTTGAATACAGTTTAGCAACAA 785
 Db 790 TTGCAAGGAAGGAAGAAATATCTCAAGTTAGCATTTGCAATACAGTTTAGCAATATAC 849
 Qy 786 GGAGATTTCGTCGGTGTGGTGGG 810
 Db 850 TGGTATTTCATCCGTAAGTGGGG 874

RESULT 11
 CK284486
 LOCUS
 DEFINITION EST747208 Nicotiana benthamiana mixed tissue cDNA library, normalized, full-length Nicotiana benthamiana cDNA clone NEMAO37 5', end, mRNA sequence.

ACCESSION
 VERSION CK284486.1 GI:39858105

KEYWORDS

SOURCE

ORGANISM

Nicotiana benthamiana
 Nicotiana benthamiana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Solanaceae; Nicotiana.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Robin Buell
 The Institute for Genomic Research
 9712 Medical Center Dr, Rockville, MD 20850, USA
 Email: potato-array@tigr.org
 Clones can be requested from the University of Arizona Genomics Institute via <http://genome.arizona.edu/orders/>.
 Seq primer: ATT TAG GTG ACA CTA TAG.

FEATURES

source

Location/Qualifiers
 1..922
 /organism="Nicotiana benthamiana"
 /mol_type="mRNA"
 /db_xref="taxon:4100"
 /clone="NEMAO37"
 /tissue_type="abiotic and biotic stress-treated leaves, callus tissue and root tissue"
 /lab_host="DH10B-TonA"
 /clone_lib="Nicotiana benthamiana mixed tissue cDNA library, normalized, full-length"
 /note="vector: pCMVSPORT6.1; Site_1: EcoRI; Site_2: NotI; supplier: RNA was isolated from Nicotiana benthamiana tissues that include callus, roots from liquid culture grown plants, heat-stressed leaves (38 C, 3 hr and 6 hr), cold-stressed leaves (5 C 3 hr, 6hr), and pathogen challenged leaves (Pseudomonas syringae pv tomato 12 hr; Xanthomonas campestris pv campestris 12 hr, 18hr; Pseudomonas syringae pv phaseolicola 18hr, and Xanthomonas campestris pv vesicatoria 18hr). RNA was isolated from these tissues and pooled in approximately equal molar amounts."

ORIGIN

Query Match 51.8%; Score 497; DB 7; Length 922;
 Best Local Similarity 75.9%; Pred. NO. 4.6e-145;
 Matches 614; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

Qy 6 GAAATAGCTTCAGCTTTGGGGAACACAGGGCTTAAGCTTAGCGCGTGGTTTGG 65
 Db 111 GACATTGAGCTCGACACTGCGGGAACACCGGCTCAACCTTAGCTCCGCGCTCGG 170

Qy 66 TGCTCTCCCTCGAAGTGTCTTCGCTCCAGTCGCGAAGATGATGTCGCTCGCACCGT 125
 Db 171 CGCTTCTCTCTTTGGCAAAAGTCTTCGCGGATGTCTCCGAAACAAGACGCTTCGCGCGCT 230
 Qy 126 GCGGAGGCTTTCCGCTCTCGGTATCAACTTCTTCGACACCTCCCGCTATTATGAGGAC 185
 Db 231 GCGGGAAGCTTTCCGCTCTCGGTCAATTCTTCTTGATCTTCCCGTATTATGAGGAC 290
 Qy 186 ACTGTCTGAGAAATGCTTGGTAAGGACTAAAGGCTTTGCAAGTCCCTAGAGTACTA 245
 Db 291 GTTATCAGAAAGGACTACTAGGAAGGCTTTGAAAGGCTCTTGGAGCTCCTAGAGATGATGA 350
 Qy 246 CATTGTGGCTACTAAGTGTGGTAGATATAAAGAAGTTTGTGATTTCACTGCTGAGAGAT 305
 Db 351 CATTGTGTCACAAAGTGTGGGAGGTACAAAGAGGGATTGTGATTTCACTGCTGAGAGAT 410
 Qy 306 AAGAAAGAGTATTGACGAGAGCTTTGGAGAGGCTTCAGCTTGTATGATTTGACATACATTTCA 365
 Db 411 GACTAAGAATTTGATGAGAGCTTTGGAGAGCTTACAGCTTGTATGATTTGATATTTTGA 470
 Qy 366 TTGCAATGACATTTGAGTTCCGGTCTCTTTGATCAGATTGTGAGTGAAACAATTCCTGCTCT 425
 Db 471 ATGTCATGATATTGAGTTTGGGTCTCTTTGATCAGATTGTGAATGAGAGCGCTTCCAGCCCT 530
 Qy 426 TCAGAAATCAAAACAGAGGGGGAAGCCCGTTTCATTTGGTATCACTGGTCTTCGTTTGA 485
 Db 531 TCAGAAATCAAGCAAGCTCGAAAGATCCGTTTTCATTTGGTATTAACCGGCTTCCATTTGGG 590
 Qy 486 TATTTTCACTTATGTTCTTCGATCGAGTCCGAGGAGCTTCGATGCTGATTTGTCATATA 545
 Db 591 GATATTCACTTATGTTGCTTGTATCGGTAACCTTCAGGAACGGTTGATGATTTCTGTCATA 650
 Qy 546 CTGTCAATTACGGGTTAATGATTTCGAGTTGCTGGATTTACTACTCTTCTTGAAGACAA 605
 Db 651 TTGTCACATCAAGTATCAATGACTCAACTTTGGAGGATCCGTTGTCATACCTGAAGAGCAA 710
 Qy 606 AGGTGGGTGTGATTAAGTGTCTTCTCAATTAGCATGAGGCTCTCTTACAGAACAGGTCC 665
 Db 711 GGGTGTGGGAGTATCATGAGTCTTCCCTCTTTCGAATGGGTCTTCTTACTGAGGCTGGAGC 770
 Qy 666 TCCTGAATGCGACCCCTGCTTCCCTGAGTCTCAAGTCTGCAAGCAAGCCGAGTTGCTCA 725
 Db 771 TCCAGATGGCACCTCTCTTCTGAACTTAAGGCTTCCCTGCGAGCTGCCGTTGATCA 830
 Qy 726 CTGCAATCAAGGCGCAAGAGATCAAAAGTTAGTCTTGCATATACAGTTTAGCAACAA 785
 Db 831 TTGCAAGGAAGGAAGAAAGATATATCAAAATTAGGCTTTGCATACAGTTTAGCAATATAC 890
 Qy 786 GGAGATTTCGTCGGTGTGGTGGGATGA 814
 Db 891 CGATATTTCATCCGTAAGTGGGGATGA 919

RESULT 12

CK286441

LOCUS

DEFINITION

EST749163 Nicotiana benthamiana mixed tissue cDNA library, normalized, full-length Nicotiana benthamiana cDNA clone NEMB212 5', end, mRNA sequence.

ACCESSION

VERSION CK286441.1

KEYWORDS

SOURCE

ORGANISM

Nicotiana benthamiana

Nicotiana benthamiana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Solanaceae; Nicotiana.

REFERENCE

AUTHORS

Buell, C.R., Hart, A., Zismann, V., Karamycheva, S.A., Day, B., Staskiewicz, B., Jin, H. and Baker, B.

Generation of EST sequences from Nicotiana benthamiana

Unpublished (2003)

Other ESTs: EST749164

COMMENT

EST 02-AUG-2004


```

Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
Clones can be requested from the University of Arizona Genomics
Institute via http://genome.arizona.edu/orders/ .
Seq primer: ATT TAG GTG ACA CTA TAG.
Location/Qualifiers
1. .896
/organism="Nicotiana benthamiana"
/mol_type="mRNA"
/db_xref="taxon:4100"
/clone="NEMB212"
/tissue_type="abiotic and biotic stress-treated leaves,
callus tissue and root tissue"
/lab_host="DH10B-Tona"
/clone_lib="Nicotiana benthamiana mixed tissue cDNA
library, normalized, full-length"
/note="vector: pCMVSPORT6.1; Site_1: EcoRI; Site_2: NotI;
supplier: RNA was isolated from Nicotiana benthamiana
tissues that include callus, roots from liquid culture
grown plants, heat-stressed leaves (38 C, 3 hr and 6 hr),
cold-stressed leaves (5 C 3 hr, 6hr), and pathogen
challenged leaves (Pseudomonas syringae pv tomatato 12 hr;
Xanthomonas campestris pv campestris 12 hr, 18hr;
Pseudomonas syringae pv phaseolicola 18hr, and Xanthomonas
campestris pv vesicatoria 18hr). RNA was isolated from
these tissues and pooled in approximately equal molar
amounts."
FEATURES
source

```

ORIGIN

Query Match	51.1%	Score 490.6	DB 7	Length 896
Best Local Similarity	77.0%	Pred. No. 4.8e-143		
Matches 598	Conservative 0	Mismatches 179	Indels 0	Gaps 0
Qy	6	GAAATAGAGCTTCGAGCTTTGGGGAACACAGGGCTTTAAGGTAGCGCCGTTGGTTTGG	65	
Db	120	GACATTGCAGCTCGACCACTGGCAACACCGGCTCAACCTTAGCTCCGTGGCTTCGG	179	
Qy	66	TGCCTCTCCGCTCGGAAGTGTCTTCGTCAGTCGCGGAAGATGATGCCCTGCCACCGT	125	
Db	180	CGCTTCTCTCTTGGCAAGTCTTTGGCGGATGTCTCGAACAAGACGCCCTTCGCGCGCT	239	
Qy	126	GCGGAGGCTTTCCGCTTCGGTATCAACTTCTTCGACACTCCCGCTATTATCGAGGAAC	185	
Db	240	GCGGAAGCCTTTCCGCTTGGCTCATTTCTTTTGATCTTCCCGCTATTATCGAGGAAC	299	
Qy	186	ACTGCTGAGAAATGCTTGGTAAGGCAATAAAGGCTTTGCAAGTCCTTAGAAGTGACTA	245	
Db	300	GTTATCAGAAAAGGTACTAGGGAAGGCTTTGAAGGCTTTGGAAGCTCCTAGAGATGAGTA	359	
Qy	246	CATTGTGCTACTAAGTGTGATAGATAAAGAGGTTTGTATTCAGTGCCTCAGAGAGT	305	
Db	360	CATTGTCTCAACAAGTGTGGAGGTTCAAGAGGGAATTTGATTTAGTCTCAGAGAGT	419	
Qy	306	AAGAAAGAGTATTGACGAGAGCTTTGGAGAGGCTTCAGCTTGATTGTTCACATACCTCA	365	
Db	420	GACTAAAGCATTCATGAGAGCTTTGGAGAGGCTACAGCTTGATTGTTCATATTGTCAT	479	
Qy	366	TTGCCATGACATTGAGTTCCGGTCTCTTGATCAGATTGTCAGTCAACAATTCCTGCTCT	425	
Db	480	ATGTCATGATTATGAGTTTGGTCTCTTGATCAGATTGTGAAGTAGAGCGCTTCACGCCCT	539	
Qy	426	TCGAAACTGAAAACAAGAGGGGAAGACCCGGTTTCATTGGTATCACTGGTCTTCGGTTAGA	485	
Db	540	TCGAAACTGAAGCAAGCTGMAAGATCCGGTTTCATTGGTATAACCGGCTTCATTGGG	599	
Qy	486	TATTTTCACTTATGTTCTTGATCGAGTGCCCTCCAGGCACTGTGATGTGATATGTCATA	545	
Db	600	GATATTCACTTATGTTGATCGGGTACCTTCAGGAAACGGTGTGATGTATTCTGTGATA	659	
Qy	546	CTGTCAATTACGGCCGTTAATGATTTCGACGTTGCTGGATTACTACCTTACCTTTGAAGAGCAA	605	

D	b		660	TTGTCACTACAGTATCAATGACTCACTTTGGAGGATCGTTGCCATCCTGAAGACAA	719
Q	y		606	AGGTGTGGGTGTGATAAGTGCTTTCTCCATTAGCAATGGGCCTCCTTTACAGAACAAGGTTC	665
D	b		720	GGGTGTGGGAGTGATCAGTGTCTTCCCCTCTCTGCAATGGGTCTTCTTACTGAGGCTGGAGC	779
Q	y		666	TCTTGAATGGCACCCCTGCTTCCCCCTGAGTCAAGTCTGCAAGCAAGAACCGCAGTTGCTCA	725
D	b		780	TCCAGAGTGGCACCCCTGCTTCTTCTGAACCTTAAGGCTGCTGCGAGCTGCCGTTGATCA	839
Q	y		726	CTGCAAAATCAAAGGCGAAGAAGATCATCAAAAGTTAGCTTGCATATACAGTTTTAGCAAA	782
D	b		840	TTGCAAGGAAGGGAAGAAGATATATCAAAATTAGCTTGCATATACAGTTTTAGCAAA	896
RESULT 13					
CK282535/c					
LOCUS					
DEFINITION					
EST745257 Nicotiana benthamiana mixed tissue cDNA library, normalized, full-length Nicotiana benthamiana cDNA clone NBMD20 3' end, mRNA sequence.					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
COMMENT					

FEATURES

```

1. .926
/organism="Nicotiana benthamiana"
/mol_type="mRNA"
/db_xref="taxon:4100"
/clone="NEMAD20"
/tissue_type="abiotic and biotic stress-treated leaves,
callus tissue and root tissue"
/lab_host="DH10B-Tona"
/clone_lib="Nicotiana benthamiana mixed tissue cDNA
library, normalized, full-length"
/notes="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
supplier: RNA was isolated from Nicotiana benthamiana
tissues that include callus, roots from liquid culture
grown plants, heat-stressed leaves (38 C, 3 hr and 6 hr),
cold-stressed leaves (5 C 3 hr, 6hr), and pathogen
challenged leaves (Pseudomonas syringae pv tomato 12 hr;
Xanthomonas campestris pv campestris 12 hr, 18hr;
Pseudomonas syringae pv phaseolicola 18hr, and Xanthomonas
campestris pv vesicatoria 18hr). RNA was isolated from
these tissues and pooled in approximately equal molar
amounts"

```

ORIGIN

	Query Match	49.98;	Score 479;	DB 7;	Length 926;
	Best Local Similarity	75.3%;	Pred. No. 2.2e-139;		
	Matches 596;	Conservative	0;	Mismatches 195;	Indels 0; Gaps 0;
Qy	170	CGTATTATGAGGAACA	CTCTCTGAGAAATGCTT	CGTAGGCACTAAAGCCTTTGCAAG	239
Db	926	CGTATATGGGAGGAA	CGTTATCAGAAAGGTA	CTAGGAGGCTTTGAGGCCTCTGGAG	867


```
Qy 230 TCCCTAGAAGTACTACATTGTGGCTAATAAGTGTGGTAGATATAAAGAGGTTTGTATT 289
Db 866 CTCCTAGAGATGAGTACATTGTGTCAACAAGTGTGGAGGTACAAAGAGGATTGTATT 807
Qy 290 TCAGTGTGAGAGAGTAAAGAGTATATGACGAGAGCTTGGAGAGGCTTCAGCTTGATT 349
Db 806 TCAGTGTGAGAGAGTGAATAAAGCATTTGATGAGAGCTTGGAGAGGCTACAGCTTGATT 747
Qy 350 ATGTTGACATCTTCAATGCCATGACATGATGATGCTGGGTCTCTTGATCAGATTCTGAGTG 409
Db 746 ATGTTGATATTTTGAATGTATGATATGATGTTGGGTCTCTTGATCAGATTCTGAGTG 687
Qy 410 AAACAATTTCTGCTCTTCAGAAACTGAAACAAGAGGGAAGACCCGGTTCATTGGTATCA 469
Db 686 AGACGCTTCCAGCCCTTCAGAACTGAGCAAGCTGGAAGATCGTTCATTGGTATNA 627
Qy 470 CTGCTTCCGTTAGATATTTTCACTTATGTTCTTGTATGAGTGCCTCCAGGAGCTGTGCG 529
Db 626 CCGGCTTCCATTTGGGATATTTCACTATGTTGCTTGTATCGGTTACTCCAGGAACGGTTG 567
Qy 530 ATGTGATATCTCATCTCATTTACGCGGTTAATGATTCGACGTTGCTGGAATTTACTAC 589
Db 566 ATGTTATTTCTGTATATTTGTCACTACAGTATCAATGACTCAACTTTGGAGGATCCGTTC 507
Qy 590 CTACTTGAAGAGCAAAAGGTGTGGGTGTGATAAGTGTCTTCTCCATTAGCAATGGCCCTCC 649
Db 506 CATACCTGAAGAGCAAGGTTGTGGAGTATGATGCTTCCCTCTTGCATGGGTCTTC 447
Qy 650 TTACAGAAACAAGGTCCTCTGAATGGCACTCTGTTCCCTCCCTGAGCTCAAGTCTCAAGCA 709
Db 446 TTACTGAGGCTGGAGCTCCAGAGTGGCACCCCTGCTTCTTCTGAACCTTAAGGCTGCCTGCC 387
Qy 710 AAGCGGAGTTGCTCACTGCAATCAAGGCAAGAGATCAAAAGTTAGCTCTGCAAT 769
Db 386 GAGTGCCTGTTGATCATTTGCAAGGAAAGGAAAGATATATCAAAATTAGCCTTGCAT 327
Qy 770 ACAGTTTAGCAAAACAAGAGATTTGCTCGGTGTGGTGGATGAGCTCTGCTCAACAGG 829
Db 326 ACAGTTTAGCAAAATACCATTTTTCATCCGTACTAGTGGGATGAATCCGTAAAGAGG 267
Qy 830 TAGAAGAAATGTTGACAGAGTTACAGAGCTTGAAGTCTCGGAGTGGATCAAGAACTC 889
Db 266 TGGAGAAATATAGGAGCTGCCCTGGAGCTAGCAACCGGCTGGGATGATGAAGAACTT 207
Qy 890 TGTCTGAGTTGAGCTATTCTCGAGCTGTAAAGATCTGACATGGCCAGTGGATCC 949
Db 206 TATCAGAGATCAGACATTTCTGAACCAAGTGAAGATCAGTCATGGCCTAGTGGTATCC 147
Qy 950 ATCAGAACTAA 960
Db 146 AACAAAGTTTA 136
```

```
RESULT 14
LOCUS CO121316
DEFINITION GR_EB02H06.f GR_Eb Gossypium raimondii cDNA clone GR_EB02H06 5',
mRNA sequence.
ACCESSION CO121316
VERSION CO121316.1 GI:48820003
KEYWORDS EST.
SOURCE Gossypium raimondii
ORGANISM Gossypium raimondii
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
1 (bases 1 to 803)
REFERENCE Kim H., Yu, Y., Kudrna, D., Hatfield, J., Stum, D., Mueller, C.,
AUTHORS Udall, J.A., Rapp, R.A., Wendel, J.F., Rao, K., Soderlund, C. and
Wing, R.A.
TITLE Global assembly of Cotton ESTs
JOURNAL Unpublished (2004)
```

COMMENT

Contact: Rod A. Wing
Arizona Genomics Institute
The University of Arizona
Forbes Building Room 303, Tucson, AZ, 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: <http://genome.arizona.edu>
Plate: 02 row: H column: 06.

FEATURES

source

1. 803

/organism="Gossypium raimondii"

/mol_type="mRNA"

/db_xref="taxon:29730"

/clone="GR_EB02H06"

/tissue_type="floral"

/dev_stage="3 to +3 DPA"

/lab_host="DH108"

/clone_lib="GR_Eb"

/notes="Vector: pCMV.SPORT-6.1; Site_1: NotI; Site_2:

EcoRV; Library made by Invitrogen with RNA supplied by

Wendle lab. Directional cloned into NotI-EV. Colonies

plated/picked by AGI. More glycerol clones held in -80."

ORIGIN

Query Match 48.8%; Score 468.6; DB 7; Length 803;

Best Local Similarity 78.4%; Pred. No. 4.1e-136;

Matches 561; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

```
Qy 5 CGAAATAGAGCTTCGAGCTTTGGGGAACACAGGGCTTAAGGTTAGCGCGCTGGTTTGG 64
Db 88 CCAACTAGAAATAGAGACCTCTCGGAAACACGGGCTCAAGCTCAGCGGTGGGCTTCG 147
Qy 65 GTGCTCTCCGCTCGGAAGTGTCTTCGCTCAGTCGCCGAAGATGATGCCGCTCCACCG 124
Db 148 GAGCTTCCCTCTCGGACAGGCTCTTCGCTTCGTTCCGAAAGCAGCGCTGCGCTCG 207
Qy 125 TGGCGAGGCTTCCGCTCGGTATCAACTCTTCGACACTCCCTCCCTATTTATGGAGAA 184
Db 208 TCCTCGAAGCTTCCGCTCGGAATCACTTCTTCGACACTCCCTCCCTATTTATGGAGCG 267
Qy 185 CACTGCTGAGAAATGCTTGGTAAGGAGCTTAAGGCTTTGCAAGTCCCTAGAGTGAAT 244
Db 268 CATTGTCAGAGAGATGCTTGGTAAGGAGCTTAAGGCTCTTGGAGTCTTGAAGAGTGA 327
Qy 245 ACATTGTGGCTACTAAGTGTGTAGATATAAGAAAGGCTTTGATTTCAGTCTCGAGAG 304
Db 328 ATATAGTTTCGACAAATGTGGAGATATCGTGAAGGTTTGTATTCAGTCTCGAGAG 387
Qy 305 TAAGAAAGATATTGACGAGAGCTTGGAGAGGCTTCAGCTTGAATATTGTTGACATATTC 364
Db 388 TAACTAAAGCATTTGATGAAGCTTGGAGAGGTTGCAACTTGAATTATGTTGATATATTC 447
Qy 365 ATTGCCATGATGATGCTTGGGTCTCTTCGATCAGATTGTGAGTGAACATTCCTGCTC 424
Db 448 AATGCCATGATGATGATGCTTGGGTCTCTTCGATCAGATTGTGAGTGAACATTCCTG 507
Qy 425 TTCAAGAACTGAAACAGAGGGAAGACCGGTTTCATTGGTATCACTGCTTCCGTTAG 484
Db 508 TTCAAAACCTGAGAGAGCAGGAGATTCGTTTCATTGATATCAACCGGTTGCGCTTG 567
Qy 485 ATATTTTCACTTATGTTCTTTGATGAGTGCCTCCAGGAGCTGTGATGATGATGATCAT 544
Db 568 AAATTTTACTTATGCTTTGATAGGCTTCACCGAGGCTGTTGATGATGATGATATATCAT 627
Qy 545 ACTGTCAATTACGGGTTTAATGATTCGAGGTTGCTGGATTACTACTTACTTGAAGAGCA 604
Db 628 ATTGCCATTAAGCATTAATGATTCACATTTGGAGGATTTATTGCTTACTTGAAGAACCA 687
Qy 605 AAGGTGGGTGTGATTAAGTGTCTTCCATTTAGCAATGGGCTCTCTACAGAACCAAGGTC 664
Db 688 AAGGTGGGTGTGATTAAGTGTCTTCCATTTAGCAATGGGCTCTCTACAGAACCAAGGTC 747
Qy 665 CTCTGAATGGCACCCCTGCTTCCCTGAGCTCAAGCTGCAAGCTGCAAGCAAGGCGCAGTT 720
```

```
Db      748 CACCGAGTGGCATCCGCGATCTCCGAACCTNAGTCTGCGCTGCCAAGCTGCTGCT 803
|||||
RESULT 15
CK297635
LOCUS      859 bp      mRNA      linear      EST 02-AUG-2004
DEFINITION Nicotiana benthamiana mixed tissue cDNA library,
normalized, full-length Nicotiana benthamiana cDNA clone NEMDA27 5',
end, mRNA sequence.
ACCESSION CK297635
VERSION   CK297635.1  GI:39884220
KEYWORDS  EST.
SOURCE    Nicotiana benthamiana
ORGANISM  Nicotiana benthamiana
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           asterids; lamids; Solanales; Solanaceae; Nicotiana.
REFERENCE Buell, C.R., Hart, A., Zismann, V., Karamycheva, S.A., Day, B.,
AUTHORS   Staskawicz, B., Jin, H., and Baker, B.
TITLE     Generation of EST sequences from Nicotiana benthamiana
JOURNAL   Unpublished (2003)
COMMENT   Contact: Robin Buell
           The Institute for Genomic Research
           9712 Medical Center Dr, Rockville, MD 20850, USA
           Email: potato-array@tigr.org
           Clones can be requested from the University of Arizona Genomics
           Institute via http://genome.arizona.edu/orders/.
           Seq primer: ATT TAG GTG ACA CTA TAG.
           Location/Qualifiers
FEATURES             source
     1..859
     /organism="Nicotiana benthamiana"
     /mol_type="mRNA"
     /db_xref="taxon:4100"
     /clone="NEMDA27"
     /tissue_type="abiotic and biotic stress-treated leaves,
     callus tissue and root tissue"
     /lab_host="DH10B-TonA"
     /clone_lib="Nicotiana benthamiana mixed tissue cDNA
     library, normalized, full-length"
     /note="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
     supplier: RNA was isolated from Nicotiana benthamiana
     tissues that include callus, roots from liquid culture
     grown plants, heat-stressed leaves (38 C, 3 hr and 6 hr),
     cold-stressed leaves (5 C 3 hr, 6hr), and pathogen
     challenged leaves (Pseudomonas syringae pv tomato 12 hr;
     Xanthomonas campestris pv campestris 12 hr, 18hr;
     Pseudomonas syringae pv phaseolicola 18hr, and Xanthomonas
     campestris pv vesicatoria 18hr). RNA was isolated from
     these tissues and pooled in approximately equal molar
     amounts."

ORIGIN
Query Match      48.2%; Score 463.2; DB 7; Length 859;
Best Local Similarity 76.6%; Pred. No. 2.1e-134;
Matches 567; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

Qy      6  GAAAAATAGAGCTTCGAGCTTTGGGGAACAAGGGTTAGCGCGCTTGGTTTGG 65
Db      114 GACATTGAGCTCCGACCACTGGGCAACACCGGCTCAACCTTAGCTCCGTCGGCTTCGG 173
Qy      66 TGCCTCTCCGTCGGAAGTGCTTCGGTCCAGTCGCGAAGATGATGCCGTGCGCACCGT 125
Db      174 CGCTTCTCTCTGGCAAGTCTTCGGCGATGCTCCGGAACAAGACGCTTCGCGCGCT 233
Qy      126 GCGCGAGCTTCCGCTCGGTATCACTTCTTCGACACCTCCCGTATTATGGAGGAC 185
Db      234 GCGCGAAGCCTTTCGCCCTTGGCGTCAATTTCTTTGATATCTCCCGTATTATGGAGAAC 293
Qy      186 ACTGTCTGAGAAATGCTTGGTAAGGACTTAAGGCTTTGCAAGTCCCTAGAAAGTACTA 245
Db      294 GTTATCAGAAAGGTACTAGGGAAGGCTTTGAGGCTCTTGAGAGCTCCTAGAGATGATA 353
```

```
Qy      246 CATTGTGGCTACTAAGTGTGTAGATATAAAGAGGTTTTGATTTCAAGTCTGAGAGAGT 305
Db      354 CATTGTGTCAACAAGTGTGGAGGTACAAAGAGGGATTTGATTTCAAGTCTGAGAGAGT 413
Qy      306 AAGAAAGAGTATTACACAGAGCTTTGGAGAGGCTTCAGCTTGATTTGTTGACATCTCA 365
Db      414 GACTAAAAGCAATTGATGAGAGCTTTGGAGAGGCTACAGCTTGATTTGTTGATATTTTGA 473
Qy      366 TTGCCATGACATTCAGATTCCGGGTCTCTTTGATCAGATTGTGAGTGAACAAATTCCTGCTCT 425
Db      474 ATGTCATGATATTGAGTTTGGGTCTCTTTGATCAGATTGTGATGAGAGCGCTTCAGGCCT 533
Qy      426 TCAGAAACTGAAACAAGAGGGGAAGACCCGGTTTCATTGGTATCACTGGTCTTCCTGTAGA 485
Db      534 TCAGAAACTGAAACAAGCTGGAAGAGATCCGGTTTCATTGGTATCAACCGGCTTCCATTGGG 593
Qy      486 TATTTTCACTTATGTTCTTTGATCGAGTGCCCTCCAGGAGCTGTGATGTGATTTGTCTATA 545
Db      594 GATATTCACTTATGTTCTTTGATCGGTTACCTCCAGGAACCGTTGATGTTATTTCTGTCTATA 653
Qy      546 CTGTCAATTACGGCGTTAATGATTCGAGTTGCTGCTGATTTTACTACCTTACTTGAAGAGCAA 605
Db      654 TTGTCACTACAGTATCAATGACTCAACTTTGGAGGATCCGTTGCGCATACCTGAAGAGCAA 713
Qy      606 AGGTGTGGGTGTGATAAGTGTCTTCCATTAGCAATGGGCTCCTTTACAGAACAAAGTTC 665
Db      714 GGGTGTGGGAGTGATCAGTGTCTTCCCTCTTTGCAATGGGTCTTCTTACTGAGGCTGGAGC 773
Qy      666 TCCTGAATGGCACCCCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAGCCGCGAGTTGCTCA 725
Db      774 TCCAGAGTGGCACCCCTGCTTCTTCTGAACTTAAGGCTGCTGCGCGAGCTGCCGTTGATCA 833
Qy      726 CTGCAAAATCAAGGGCAAGA 745
Db      834 TTGCAAAAGGAAGGGAAGA 853
```

Search completed: November 13, 2005, 08:03:38
Job time : 3739 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 13, 2005, 08:07:21 ; Search time 23 Seconds
(without alignments)
6231.574 Million cell updates/sec

Title: US-10-606-300-12
Perfect score: 1653
Sequence: 1 atgacgaaatagagcttcg.....gtggaatccatcagaactaa 960

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_b/US10606300/runat_07112005_094820_7323/app_query.fasta_1.1159
-DB=Issued_Patents_AA -QFMT=fascan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10606300@cgn 1 1 22 @runat_07112005_094820_7323 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PTCUS.COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1619	97.9	319	4	US-09-630-983A-11
2	317	19.2	367	4	US-09-252-991A-26618
3	314.5	19.0	347	4	US-09-902-540-10503
4	302.5	18.3	341	4	US-09-252-991A-30076
5	300.5	18.2	332	4	US-09-902-540-10805
6	296.5	17.9	332	4	US-09-107-532A-6891
7	296	17.9	334	4	US-09-328-352-6079
8	291.5	17.6	314	3	US-09-134-001C-3126
9	291	17.6	348	4	US-09-673-198-4
10	277.5	16.8	364	4	US-09-489-039A-12541
11	277	16.8	334	4	US-09-724-623-86
12	274	16.6	350	4	US-09-489-039A-9552

13	271	16.4	328	4	US-09-489-039A-12803	Sequence 12803, A
14	264	16.0	358	4	US-09-252-991A-31535	Sequence 31535, A
15	254	15.4	347	4	US-09-489-039A-9489	Sequence 9489, Ap
16	252.5	15.3	325	4	US-09-902-540-12689	Sequence 12689, A
17	250	15.1	367	4	US-09-949-016-6399	Sequence 6399, Ap
18	249	15.1	324	4	US-09-949-016-10971	Sequence 10971, A
19	247	14.9	353	4	US-09-328-352-6721	Sequence 6721, Ap
20	245.5	14.9	314	3	US-09-134-001C-5405	Sequence 5405, Ap
21	245	14.8	329	2	US-08-606-143-2	Sequence 2, Appl
22	244.5	14.8	351	4	US-09-252-991A-19370	Sequence 19370, A
23	242	14.6	329	2	US-08-606-143-1	Sequence 1, Appl
24	242	14.6	329	2	US-08-606-143-3	Sequence 3, Appl
25	234.5	14.2	319	4	US-09-902-540-11071	Sequence 11071, A
26	233	14.1	314	4	US-09-902-540-14628	Sequence 14628, A
27	221	13.4	330	4	US-09-679-279-10	Sequence 10, Appl
28	219.5	13.3	290	4	US-09-602-787A-670	Sequence 670, App
29	219.5	13.3	380	4	US-09-248-796A-17336	Sequence 17336, A
30	218.5	13.2	316	4	US-09-134-000C-5598	Sequence 5598, Ap
31	218.5	13.2	317	4	US-09-252-991A-22776	Sequence 22776, A
32	218	13.2	357	4	US-09-252-991A-24365	Sequence 24365, A
33	217.5	13.2	353	4	US-10-092-263-8	Sequence 8, Appl
34	217	13.1	1114	2	US-08-576-626A-31	Sequence 31, Appl
35	216	13.1	342	4	US-09-949-016-7654	Sequence 7654, Ap
36	214.5	13.0	310	4	US-09-252-991A-32639	Sequence 32639, A
37	214	12.9	306	4	US-09-902-540-9859	Sequence 9859, Ap
38	214	12.9	342	4	US-09-538-092-776	Sequence 776, App
39	213.5	12.9	353	4	US-09-248-796A-17342	Sequence 17342, A
40	211	12.8	398	4	US-09-489-039A-9541	Sequence 9541, Ap
41	209	12.6	330	4	US-09-724-797-28	Sequence 28, Appl
42	208.5	12.6	291	4	US-09-107-532A-5063	Sequence 5063, Ap
43	201.5	12.2	255	4	US-09-583-110-3418	Sequence 3418, Ap
44	200.5	12.1	321	4	US-09-489-039A-13361	Sequence 13361, A
45	200	12.1	366	4	US-09-248-796A-17341	Sequence 17341, A

ALIGNMENTS

RESULT 1

US-09-630-983A-11
; Sequence 11, Application US/09630983A
; Patent No. 6630330
; GENERAL INFORMATION:
; APPLICANT: Porro, Danilo
; APPLICANT: Sauer, Michael
; TITLE OF INVENTION: Ascorbic Acid Production from Yeast
; FILE REFERENCE: 2028.594000
; CURRENT APPLICATION NUMBER: US/09/630.983A
; CURRENT FILING DATE: 2000-08-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-630-983A-11

Alignment Scores:		Pred. No.:	1.7e-169	Length:	319
Score:	1619.00	Matches:	319		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	97.94%	Indels:	0		
DB:	4	Gaps:	0		
US-10-606-300-12 (1-960) x US-09-630-983A-11 (1-319)					
Qy	1	ATGACGAAATAGAGCTTCGAGCTTTGGGGAACACAGGGCTTAAGTTAGCGCCCTTGGT	60		
Db	1	MetThrIysIleGluLeuArgAlaLeuGlyAsnThrGlyLeuIysValSerAlaValGly	20		
Qy	61	TTTCGGTCCTTCGGCTCGGAGGTCTTCGGTCCAGTCGCGGAGATGATGCGCTCGCC	120		
Db	21	PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAla	40		

```
QY 121 ACCGTCGCGAGGCTTTCCGTCTCGGTATCAACTTCTTCGACACCTCCCGTATATATGGA 180
Db 41 ThrValArgGluAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyrTyrGly 60
QY 181 GGACACTGCTGAGAAAATCGTTGGTAAGGACTAAAGGCTTTCGAAGTCCCTAGAGT 240
Db 61 GlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSer 80
QY 241 GACTACATTGGCTTACTAAGTGTGGTATAGATATAAAGAGGTTTGTGATTTCAAGTCTCAG 300
Db 81 AspTyrIleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGlu 100
QY 301 AGAGTAAGAAAGATGATTGACAGAGCTTGGAGAGCTTCAGCTTGATTTGTTGACATA 360
Db 101 ArgValAlaGlySerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 120
QY 361 CTTCAATGCCCATGACATTGAGTTCGGGTCTCTTGATCAGATTGTGAGTGAACATTTCT 420
Db 121 LeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePro 140
QY 421 GCTCTTCAGAACTCAAAACAGAGGGAAGACCCGTTTCATTGGTATCACTGGTCTTCGG 480
Db 141 AlaLeuGlnLysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGlyLeuPro 160
QY 481 TTAGATATTTCACTTATGTTCTTCATCGAGTGCCTCCAGGAGCTGTCCGATGTGATATG 540
Db 161 LeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeu 180
QY 541 TCATACCTCATAGCGGCTTAATGATTCGAGGTTCGCTGGATTTACTACCTACTTCTGAAG 600
Db 181 SerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLys 200
QY 601 AGCAAGGTGTGGGTGTGATAGTGTCTTCCATTAGCAATGGGCTCTTTCACAGAACAA 660
Db 201 SerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGln 220
QY 661 GGTCTCTCTGAATGACGCTGCTTCCCTGAGCTCAAGTGTGCAAGCAAGCCGAGTT 720
Db 221 GlyProProGluTrpHisProAlaSerProGluLeuLysSerAlaSerLysAlaAlaVal 240
QY 721 GCTCACTCCCAATCAAGGCAAGAGATCAAAAGTTAGCTCTGCAATACAGATTTAGCA 780
Db 241 AlaHisCysLysSerLysGlyLysLysIleThrLysLeuAlaLeuGlnTyrSerLeuAla 260
QY 781 AACAAAGAGATTTCTGTCGGTGTGGTGGATGAGCTCTGCTCACAGGTAGAGAAAAT 840
Db 261 AsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGln 280
QY 841 GTTCAGCAGATTACAGAGCTTCAAGTGTGGGATGGATCAAGAACTCTGTCTCAGGTT 900
Db 281 ValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluVal 300
QY 901 GAAGTATTTCTGAGCCTGTAAGATCTGACATGCCCAAGTGGGAATCCATCAGAAC 957
Db 301 GluAlaIleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGlnAsn 319
```

RESULT 2

```
US-09-252-991A-26618
; Sequence 26618, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
```

```
; SEQ ID NO 26618
; LENGTH: 367
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26618
```

```
Alignment Scores:
Pred. No.: 6,22e-26 Length: 367
Score: 317.00 Matches: 84
Percent Similarity: 50.17% Conservative: 60
Best Local Similarity: 29.27% Mismatches: 103
Query Match: 19.18% Indels: 40
Db: 4 Gaps: 8
```

US-10-606-300-12 (1-960) x US-09-252-991A-26618 (1-367)

```
QY 55 GTTGGTTTGTGCTCTCCGTCGGAAGTGTCTTCGGTCCAGTCGCCGGAAGATGATGCC 114
Db 52 IleGlyLeuGlyGlyAlaProLeuGlyAsnMetPheHisProLeuSerGluGluThrAla 71
QY 115 GTCGCCACCGTCGCGAGGCTTCCGTCTCGGTATCAACTTCTTCGACACCTCCCGTAT 174
Db 72 AspAlaThrLeuAsnAlaAlaTrpAspAlaGlyPheArgTyrTyrAspValSerProHis 91
QY 175 TATCGAGAACACTGTCTGAGAAAATGCTTGGTAAGGACTAAAGGCTTTCGAAGTCCCT 234
Db 92 TyrGlyAlaGlyLeuAlaGluGlnArgPheGlyArgLeuLeuSerGly-----LysPro 109
QY 235 AGAAGTGTACATTTGTGGCTACTAAGTGTGCTAGATAT----- 273
Db 110 ArgAspGluTyrValLeuSerThrLysValGlyArgLeuLeuGlnProAlaSerGlnPro 129
QY 274 ----- 297
Db 130 GluAsnAlaLysProPheValAspGluLeuProAsnLysArgValProAspTyrSerAla 149
QY 298 GAGAGAGTAAAGAGAGTATTGACGAGAGCTTGGAGAGCTTCAGCTTGATTTATTTGAC 357
Db 150 AspGlyAlaArgSerIleGluAspSerLeuGluArgMetGlyValAspArgLeuAsp 169
QY 358 ATACTTTCATTCGCATGACATT-----GAGTTCGGGTCTCT-----CTT 393
Db 170 ValValPheIleHisAspValSerGluAspGlnTrpGlyProGlnTrpArgGluTyrPhe 189
QY 394 GATCAGATTGTGAGTAAACAAATTCCTCTCTTCGAAACCTGAAACAGAGGGGAGACC 453
Db 190 GlnGlnAlaMetAsnGlyAlaAlaLysAlaLeuThrGlnLeuArgAspGluGlyValIle 209
QY 454 CGGTTTCATTGATATCAGTGGTCTTCGGTTAGATATTTTTCACCTTATGTTCTTGATCGAGTG 513
Db 210 ArgGlyTrpGlyLeuGlyValAsnLeuValGluProCysArgLeuAlaLeuGluGlnSer 229
QY 514 CCTCAGGAGCTGTGATGTGATATTGTCTATCTACTTACCTACCGGCTT-----AATGAT 567
Db 230 AspPro-----AsnValPheLeuLeuAlaGlyArgTyrSerLeuLeuGluHisAsp 246
QY 568 TCGAGTTGTGCGATTATTACTACCTTACTTGAAGACAAAGGTGTGGGTGTGATAAGTGTCT 627
Db 247 GluAlaLeuAspThrLeuPheProThrCysGlnAlaArgAspValGlyValValGly 266
QY 628 TCTCCATTAGCAATGGGCTCTCTTACAGAACAAAGTCTCTCGAATGACCACTGCTTCC 687
Db 267 GlyProPheAsnSerGlyValLeuAlaGlyAspHisTyrGlyTyrAspGlnIlePro 286
QY 688 CTGAGCTC-----AAGTCTCAAGCAAGCCGAGTTGTCTACTCTGCAAAATCAAG 738
Db 287 ProGlnValAlaGlnArgArgGluGlnLeuLysAlaAlaGluHisCys----- 303
QY 739 GCGAAGAAGATCACAAAGTTAGCTCTGCAATACAGTTTACAGCAAGAGAGATTTCTGTCG 798
Db 304 GlyValAspLeuArgAlaAlaLeuHisPheCysLeuAlaAsnProValValAlaSer 323
QY 799 GTGTTGGTTGGATGAGCTCT 819
```

```
Db 324 ValIleProGlyThrAlaAsn 330
US-09-902-540-10503
; Sequence 10503, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 10503
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-10503

Alignment Scores:
Pred. No.: 1.14e-25 Length: 347
Score: 314.50 Matches: 87
Percent Similarity: 46.13% Conservative: 62
Best Local Similarity: 26.93% Mismatches: 119
Query Match: 19.03% Indels: 55
DB: 4 Gaps: 9

US-10-606-300-12 (1-960) x US-09-902-540-10503 (1-347)
Qy 10 ATAGAGTTCAGCTTGGGAAACACAGGCTTAAGTTAGCGCGTGGTTTGGTCCC 69
Db 12 ValProThrAsnAlaProGlyAsnLysPheArgProThrThrArgIleGlyLeuGly 31
Qy 70 TCTCCGTCGGAAGTCTTCGGTCCAGTCGCGGAAGATGATGCGTCGCCACCGTGGC 129
Db 32 ValAlaIleGlyAsnGlyPheAlaSerThrSerPheAlaGlnAlaGlnAlaThrLeuGlu 51
Qy 130 GAGGCTTTCGCTCGGTATCAACTTCTTGACACCTCCCGGTATATGAGGAACTG 189
Db 52 AlaAlaTAlaAlaGlyValArgTyrPheAspThrSerProTyrGlyLeuGlyLeu 71
Qy 190 TCTGAGAAATGCTTGGTAAGGACTAAAGGCTTTCGCAAGTCCCTAGAGTCACTATT 249
Db 72 SerGluArgArgPheGlyArgPheLeuHisThrArgAspAla-----AsnGluTyrVal 89
Qy 250 GTGGCTACTAAGTGGTAGA----- 270
Db 90 LeuSerThrLysValGlyArgIleLeuThrAlaAlaAspAlaProProLysLeuLeu 109
Qy 271 -----TATAAAGAGGTTTTCATTTTCAGTGTCTGAGAGAGTAAAGAAAG 312
Db 110 TrpSerAspAlaSerProPheAsnTyrGlnTyrAspTyrSerAlaAlaGlyValArgArg 129
Qy 313 AGTATTGACGAGAGTTCGAGAGGCTTCAGCTTGATTGATTGACATCTTCATTCGCAT 372
Db 130 SerValGluAspSerLeuGlnArgLeuGlyValSerArgIleAspIleValPheIleHis 149
Qy 373 GACATT-----GAGTTCGGGTCT-----CTTGATCAGATT 402
Db 150 AspLeuSerProAspAsnGlnAspMetGlyAspLysTyrThrGluTyrPheAspGlnAla 169
Qy 403 GTGAGTGAACAATTCCTGCTCTTCAGAAACTGAAACAGAGGGGAGAACCCGGTTCATT 462
Db 170 IleLysGlyAlaIleProGluLeuThrLysMetArgGluGluGlyLeuIleLysAlaTrp 189
Qy 463 GGTATCACTGGTCTTCGTTAGATATTTCTACTTANGTCTTGTGATCGAGTCGCTCCA--- 519
Db 119

us-10-606-300-12.n2p.ra1
Db 190 Gly-----PheGlyValAsnArgAlaGluProAla 199
Qy 520 -----GGGACTGTGATGATTGTCTCATCTGCTATTACGGCGTT 561
Db 200 LeuArgAlaLeuGluValSerAspProAspIlePheLeuLeuAlaThrGluTyrSerLeu 219
Qy 562 -----AATGATTGACGCTGCTGCTGATTACTTACTTCTTGAAGAGCAAGGTGTGGT 615
Db 220 ThrAsnHisGluGluAlaLeuSerLysThrPheProAlaLeuGluLysArgGlyIleSer 239
Qy 616 GTGATAAGTCTCTCCATTAGCAATGGCCCTCTTACAGAACCAAGGCTCCTCCTCAATGG 675
Db 240 ValValValGlySerProLeuAspAlaGlyTyrLeuAlaGlyArgAsnArgTyrLeuTyr 259
Qy 676 CAC-----CCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAGCGCGAGTTGCTCAC 726
Db 260 AspGlyThrIleProGlyValGlnGlyLysArgAlaArgMetSerAlaIleAla--- 278
Qy 727 TGCAAATCAAAGGCGAAGAGATCAAAAGTTAGTCTTGCATATACAGTTTACCAACCAAG 786
Db 279 ---AlaLysHisGlyIleAspLeuArgThrAlaAlaLeuGlnPheAlaAlaProSer 297
Qy 787 GAGATTTTCGTCGCTGTTGGTGGATGAGCTCTCTCACAGGTAGAGAAATCTTCCA 846
Db 298 ValValSerSerValIleProGlyAlaArgThrProGluGlnValLysAlaAsnValGln 317
Qy 847 GCAGTTACA 855
Db 318 SerMetThr 320

RESULT 4
US-09-252-991A-30076
; Sequence 30076, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30076
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30076

Alignment Scores:
Pred. No.: 2.37e-24 Length: 341
Score: 302.50 Matches: 94
Percent Similarity: 46.63% Conservative: 58
Best Local Similarity: 28.83% Mismatches: 127
Query Match: 18.30% Indels: 47
DB: 4 Gaps: 11

US-10-606-300-12 (1-960) x US-09-252-991A-30076 (1-341)
Qy 25 TTGGGGAACACAGCGCTTAAGTTAGCGCCGCTGTTGGTGGCTCTCCGCTCGGAAGT 84
Db 24 LeuGlyAsnSerGlyLeuLysValSerArgLeuGlyCysMetThrTyrGlyAsp 43
Qy 85 ---GTCTTCGGTCCA-----GTCCCGAAGATGATGCCCGCCACCGTGGCGAGGCT 135
Db 44 ProAlaTrpArgProTyrValLeuAspGluArgAlaArgProPheIleArgGluAla 63
Qy 136 TTCCGCTCGGTATCAACTTCTTGCACACCTCCCGTATTTATGGAGGAACACTGTCTGAG 195
Db 64 LeuGluAlaGlyIleAspPhePheAspSerAlaAspIleTyrSerThrGlyGluSerGlu 83
```

```
QY 196 AAAATGCTGGTAAAGGACTTAAAGGCTTTGCAAGTCCCTAGAGTGACTACATTTGGCT 255
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 ArgIleLeuGlyArgAlaLeuArgAspPhe---AlaGlnArgGluAspLeuValIleAla 102
QY 256 ACTAAG-----TGTTGGTAGATATAAAGAGGT 282
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103 ThrLysAlaPhePheProMetSerAspArgProAsnAlaCysGly----- 117
QY 283 TTTGATTTTCAGTCTGAGAGACTAAGAAGAGTATTGACGAGCTTCGAGAGGCTTCAG 342
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
118 -----LeuSerArgLysHisLeuLeuAlaSerValAspAlaSerLeuArgLeuGly 135
QY 343 CTTGATTATGTTGACATCTTCATTGCGCATGACATTGAGTTGCGGTCTCTTCATCAGATT 402
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
136 ThrAspTyrLeuAspLeuPheValIleHisArgPheAsp-----ProGluThrPro 152
QY 403 GTGAGTGAACAACTTCGCTCTTCAGAACTGAAACAGAGGGGAAAGCCCGGTTCAATT 462
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153 IleGluGluThrCysGluThrLeuAspSerLeuValArgAlaGlyLysValArgTyrLeu 172
QY 463 GGTATCACTGGTCTCCGTTAGATATTTCACTTATGTTCTTGAT---CGAGTGCCTCCA 519
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
173 GlyAlaSerSerMetProAlaTrpArgPheMetLysMetLeuAlaPheGlnArgHis 192
QY 520 GGGACTGTGATGTGATATTGTCATCTGTCATTACGGCGTT-----AATGATTGC 570
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
193 GlyLeuAlaGlnPheIleSerMetGlnSerGlnTyrAsnLeuIleValArgGluAspGlu 212
QY 571 ACGTTGCTGGATTACTACTTACTTGAAGAGCAAGGTGTGGGTGTGATGATGCTTCT 630
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
213 Glu-----AspLeuValProLeuCysArgGluGluGlyIleAlaLeuThrProTrpSer 230
QY 631 CCATTAGCAATGGGCTCCTT-----ACA 654
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
231 ProLeuAlaArgGlyLeuLeuAlaGlyAlaArgSerAlaGlyThrLeuArgThrArgThr 250
QY 655 GAACAAGGTCTCTGAAATGGCACCCCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAGCC 714
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251 AspGluGlnAlaProArgTrpTyrGlyArgGluGluValGluSerThrLeuGlyAla 270
QY 715 CGAGTTGCTCACTGCAATCAAGGGCAGAGATCACAAAGTTAGCTCTGCAATACAGT 774
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
271 LeuGluLysLeuAlaAlaArgGlyLeuProProAlaGlnLeuAlaLeuAlaTrpLeu 290
QY 775 TTAGCAAAACAAGGAGATTTCGCGGTGTGGTGGATGAGTCTGTCTCACAGGTAGAA 834
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
291 LeuGlyArgAsnGlyValAlaAlaProIleValGlyLeuSerArgProHisLeuGlu 310
QY 835 GAAATGTTGACGAGTTACAGAGCTTGAAGTCTGGGATGGATCAAGAACTCTGTCT 894
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
311 AspAlaLeuAlaAlaLeuThr-----LeuAspLeuAlaGluGluCysAla 326
QY 895 GAGTTGAGCTATTCTC 912
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
327 ThrLeuGluAlaProLeu 332
```

RESULT 5

```
US-09-902-540-10805
; Sequence 10805, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
```

```
; SEQ ID NO 10805
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-10805
```

Alignment Scores:

```
Score: 3.88e-24 Length: 332
      300.50 Matches: 97
Percent Similarity: 46.57% Conservative: 59
Best Local Similarity: 28.96% Mismatches: 128
Query Match: 18.18% Indels: 51
      4 Gaps: 11

US-10-606-300-12 (1-960) x US-09-902-540-10805 (1-332)

QY 10 ATAGAGCTTCAGCTTTGGGGAACACAGGGCTTAAAGTTAGCGCGTGGTTTGGTGCC 69
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MetHisPheArgLysLeuGlyArgSerGlyLeuValSerGluIleSerTyrGlyAsn 20
QY 70 ---TCTCCGCTCGGAAGTGTCTTCGGTCCACTGCCGCGAAGATGTCGTCGCCACCGTG 126
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
21 TrpIleThrHisGlySer-----GlnValGluGluAlaLeuAlaCysVal 37
QY 127 CGCAGGCTTCCGCTCCGCTATCACTTTCGACACCTCCCGCTATTATGGAGAAC 186
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
38 ArgAlaAlaLeuAspValGlyIleThrThrPheAspThrAlaAspValTyrAlaAlaThr 57
QY 187 CTGCTGAGAAAATGCTTGGTAAGGGACTAAAGGCTTTGCAAGTCCCTAGAGTCACTAC 246
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
58 ArgAlaGluGluValLeuGlyArgAlaLeuLysGlyGluArg-----ArgAlaGlyTyr 75
QY 247 ATTGTGGCTACTAAG-----TGTGCTAGATATAAAGAGGTTTGTGATTTCACT 294
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
76 GluLeuPheThrLysValTyrTrpProThrGlyProGlyLysAsnAspArgGlyLeuSer 95
QY 295 GCTGAGAGAGTAAGAAGAGTATTGACGAGAGCTTGGAGAGCTTCAGCTTGATTTGTT 354
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96 ArgLysHisIleLeuGluSerIleAspGlySerLeuArgArgLeuGlnThrAspTyrVal 115
QY 355 GACATACCTTCATTCGATGACATTCAGTTTCGGTCTCTTGATCAGATGTGAGTCAAAACA 414
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116 AspLeuTyrGlnAlaHisArgPheAsp-----ValGluThrProLeuGluGluThr 132
QY 415 ATTCTGCTCTTCAGAAACTGAAACAGAGGGGAGAGCCCGGTTTCATTGGTATCACTGT 474
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
133 MetLeuAlaPheAlaAspIleValArgGlnGlyLysAlaLeuTyrIleGlyValSer--- 151
QY 475 CTTCGGTTAGATATTTTCACCTTATCTTCTTCATCGAGTGCCTCCAGGGACT----- 525
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
152 -----GluTrpThrAlaAspGlnIleArgGlnGlyAlaAlaLeuAla 165
QY 526 -----GTCGATGTGATATTGTCATCTGTCATTACCGGCTTAATGATTTCGAGC 573
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
166 ArgGluLeuArgValProPheIleSerAsnGlnProGlnTyrSerMetLeuTyrArgVal 185
QY 574 TTG---CTGATTACTACTTACTTGAAGAGCAAGAGTGTGGGTGTGATAGTCTTCT 630
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
186 IleGluProGlnValIleProAlaSerAspGluAlaGlyLeuGlyGlnIleValTrpSer 205
QY 631 CCATTAGCAATGGGCTCTCTTACA-----GAACAGGTCCTCTGTAATGG 675
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
206 ProfileAlaGlnGlyValLeuThrGlyLysTyrLeuProGlyGlnAlaProProAlaGly 225
QY 676 CACCCTGCT-----TCCCTGAG 693
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
226 SerArgAlaThrGluAlaAsnAlaValArgTyrGlyIleThrArgPheMetThrAspAsp 245
QY 694 CTCAGTCTGCAAGCAAGCCGCGAGTGTCTCACTCAATAATCAAGGGCAAGAGATCACA 753
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
246 ValLeuThrArgValGlnGlnLeuValProLeuAlaLysAspValIleLysSerMetAla 265
QY 754 AAGTTAGCTCTGCAATACAGTTTAGCAAAAGAGGATTTCTCGTGGTGTGGTGGATG 813
```

Db 266 GlnLeuAlaValAlaTrrpValLeuGlnAenProSerValSerAlaAlaValGlyAla 285
QY AGCTCTGCTCACAGTAAAGAAATGTTGCAGCAGTTACAGAGCTTGAAGAGTCTGGGG 873
Db 286 SerArgProGluGlnValHisAspValValysAlaAla-----GlyValys 301
QY 874 ATGATCAAGAACTCTGCTGAGGTTGAAGCTATTCTCGAGCCT 918
Db 302 LeuGluProGluLeuLeuArgArgileAspAlaValLeuGlyPro 316

RESULT 6

US-09-107-532A-6891

; Sequence 6891, Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; CORRESPONDENCE ADDRESS: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

; ADDRESS: GENOME THERAPEUTICS CORPORATION

; STREET: 100 Beaver Street

; CITY: Waltham

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02354

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD-ROM ISO9660

; COMPUTER: PC

; OPERATING SYSTEM: <Unknown>

; SOFTWARE: ASCII

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/107,532A

; FILING DATE: 30-Jun-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/085,598

; FILING DATE: 14 May 1998

; APPLICATION NUMBER: 60/051571

; FILING DATE: July 2, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Ariniello, Pamela Deneka

; REGISTRATION NUMBER: 40,489

; REFERENCE/DOCKET NUMBER: GTC-012

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (781)893-5007

; TELEFAX: (781)893-8277

; INFORMATION FOR SEQ ID NO: 6891:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 332 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: YES

; ORIGINAL SOURCE:

; ORGANISM: Enterococcus faecium

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (B) LOCATION 1...332

; SEQUENCE DESCRIPTION: SEQ ID NO: 6891:

US-09-107-532A-6891

Alignment Scores:

Pred. No.:	1,07e-23	Length:	332
Score:	296.50	Matches:	88
Percent Similarity:	47.19%	Conservative:	63
Best Local Similarity:	27.50%	Mismatches:	132
Query Match:	17.94%	Indels:	37
DB:	4	Gaps:	10

US-10-606-300-12 (1-960) x US-09-107-532A-6891 (1-332)

QY 13 GAGCTTCGAGCTTTGGGGAACACAGGCTTAAGGTAGCGCGTTGGTTTGGTGGCTCT 72

Db 20 GluIleArg-----IleGlyHisSerGlnValTyrAlaGluGlnLeuGlyLeuGlyAlaAen 38
QY CGCTCGGA-----AGTGCTCTCGTCCAGTCCGCGAAGATGATGCGCGTCCGCCACCGTG 126
Db 39 AlaValGlyGlyHisAsnLeuPheAspGlyLeuGluAspGluThrGlyValValVal 58
QY 127 CGCAGGCTTTCCGCTCGGTATCAACCTTTCGACACCTCCCGGTATATTAGGAGAACAA 186
Db 59 ArgThrAlaLeuAenSerGlyIleAsnLeuIleAspThrAlaTyrAlaTyrGlyAenGly 78
QY 187 CTGCTCGAGAAATGCTTGGTAAGGCTTAAGGCTTTCGCAAGTCCCTAGAGTCACTAC 246
Db 79 ArgSerGluGluLeuIleGlyValLeuLysGluLysGluLysGluLysArgSerArgVal 98
QY 247 ATTGTGCTACTAAGTGTGTAGATATAA-----GAAGGTTTGTATTTCAGTGTCT 297
Db 99 ValIleAlaThrLysAlaAlaHisValProAsnLysGlyArgThrPheAspAsnSerPro 118
QY 298 GAGAGAGTAAGAAAGATATTGACGAGAGCTTGAGAGCTTCAGCTTGATTTATTTGAC 357
Db 119 GluPheLeuLysGlnSerValGluAspAlaLeuLysArgLeuGlnThrAspTyrIleAsp 138
QY 358 ATACTTCATTCCTCCATGACATTGAGTTCGGTCTCTTGATCAGATTGTGAGTGAACAATT 417
Db 139 IlePheTyrIleHis-----PheProAspGluSerThrProLysAsnGluSerVal 155
QY 418 CTGCTCTCTTCAAGAACTGAAAGAGGGGAGACCGGTTTCATTGGTATCACTGGTCTT 477
Db 156 AlaThrLeuHisGluLeuLysGluAlaGlyLysIleArgAlaValGlyValSerAsnPhe 175
QY 478 CCGTTAGATATTTCATTTATGTTCTTGATCGAGTGCCTCCAGGAGCTGTGATGTGATA 537
Db 176 ThrLeuGlu-----GlnLeuLysGluAlaAsnAlaAspGlyTyrValAspValVal 192
QY 538 -----TTGTCTACTCTGCTTACGCGCTTAATGATTTCGACGTTCTGATTTACTA 588
Db 193 GluAspLysTyrSerLeuIleHisArgGlnAlaGluLys-----GluLeuPhe 208
QY 589 CTTTACTTGAAGAGCAAGAGTGTGGTGTAGTAGTGTCTCTCCATTAGCAATGGCGCTC 648
Db 209 ProTyrLeuGluLysAsnLysIleSerPheValProTyrPheProLeuAlaSerGlyLeu 228
QY 649 CTTACA-----GAACAAGTCTCTCTGAAATGGCACCCCT 681
Db 229 LeuThrGlyLysTyrGluLeuGlyGluLysGlnPheGlyGluGlyAspProArgLys 248
QY 682 CTTCCCTGAGCTCAAGTCTGCAAGCAAGCCGAGTGTCTCACTGCAAAATCAAAGGCG 741
Db 249 ArgAsnProAspPheGlnGlyGluArgPheArgGluIleLeuThrAlaValAspValLeu 268
QY 742 AAGAAGATCACAAG-----TTAGCTCTGCAATACAGTTTA 777
Db 269 ArgProIleAlaLysArgTyrGlnAlaThrProAlaGlnLeuValLeuAlaTrrpTyrMet 288
QY 778 GCAAAACAGGAGATTTCCGTCGGTGTGGTGGATGAGCTCTGTCTCACAGTAGAAGAA 837
Db 289 LysAsnProArgValSerValIleProGlyAlaLysArgProGluGlnValSerAsp 308
QY 838 AATGTTGACGAGTTCAGAGCTTGAAGTCTGGGATGGATCAAGAAGTCTGTCTGAG 897
Db 309 AsnValGlnAlaLeu---AspLeuHisLeuSerAsnGluAspTyrGlnThrIleAspGlu 327

RESULT 7

US-09-328-352-6079

; Sequence 6079, Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

```
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6079
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6079

Alignment Scores:
Pred. No.: 1,22e-23 Length: 334
Score: 296.00 Matches: 93
Percent Similarity: 44.93% Conservative: 62
Best Local Similarity: 26.96% Mismatches: 128
Query Match: 17.91% Indels: 62
DB: 4 Gaps: 10

US-10-606-300-12 (1-960) x US-09-328-352-6079 (1-334)

QY 7 AAAATAGAGCTTCGAGCTTTGGGGAACACAGGGCTTAAGTTAGCGCGCTGTTGGT 66
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1 LysMetLysTyrArgThrLeuGlyGlnThrGlyGluLysValSerAlaLeuGlyLeuGly 20
QY 67 GCCTCCGCTCGGAAGTCTCTCCGTCAGTCCAGTCGCGGAAGATGATGCCGTCACCGTG 126
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
21 CysMetGlyMetSerPheAlaTyrGlyAlaSerAspThrGlnSerIleAlaThrLeu 40
QY 127 CCGGAGGCTTCCGCTCGGTATCAACTTCTTCGCACACCTCCCGCTATTATGGAGGAACA 186
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
41 GluLysAlaLeuAspLeuGlyIleAsnPheTyrAspThrAlaAspMetTyrGlyAsnGly 60
QY 187 CTGCTCGAGAAATGCTTGGTAAGGACTAAAGGCTTTCGAAGTCCCTAGAGTCACTAC 246
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
61 AlaAsnGluValLeuSer-----LysValLeuGluLysHisArgAspLysVal 77
QY 247 ATTGGGCTACTAAGTGTGT---AGATATAAGAGAGT----- 282
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
78 PheLeuAlaThrLysPheGlyPheArgTyrLysGluAspAsnLeuAsnProLysAsnSer 97
QY 283 -----TTTGATTTCAGTCTGAGAGAGTAAGAAAGATATTGACGAGAGCTTG 330
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
98 LeuGluSerTyrIleAspGlySerProGluTyrIleLysValAlaValGluAsnSerLeu 117
QY 331 GAGAGCTTCAGCTGATTATGTGACATCTTCACTTCATGTCATGACATGAGTTCGGTCT 390
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
118 ArgArgLeuAsnThrAspValIleAspLeuTyrTyrAlaHisArgIleAsp----- 134
QY 391 CTTGATCAGATTGTGAGTGAACAATCTCTGCTTCAGAAACTGAAACAGAGGGAAG 450
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
135 ProAsnValProValGluAspThrIleGlyAlaMetAlaAspLeuValLysGlnGlyLys 154
QY 451 ACCGGTTCATTGGTATCATCTGCTTCGTTGATATTTTCACCTATTATGTTCTTGATCGA 510
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155 ValArgTyrLeuGlyLeuSerGluAlaSerAlaGluThrIleArgLysAlaHisAlaIle 174
QY 511 GTGCTCCAGGAGCTGTCATGTGATATTGTCTACTGTCATACGGCGTTAATGATTCG 570
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
175 HisProIleAlaAlaValGln-----HisGluTyr 184
QY 571 AGTTGCTGGATTTA-----CTACCTTACTTGAAGCAAGCAAGGTG 612
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
185 SerLeuThrArgGluPheGluGlnThrHisLeuGlnThrIleArgGluLeuGlyIle 204
QY 613 GGTGTGATAAGTCTTCATTGCAATGGCGCTCTTACAGAA----- 657
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
205 SerLeuValProTyrSerProLeuSerArgGlyLeuIleThrAsnThrLeuAspValAsn 224
QY 658 -----CAAGTCCTCCTGAATGG 675
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
225 AsnLeuAspGluAsnAspPheArgGlnLeuProArgTyrGlnGly---AspAsnTrp 243
QY 676 CACCCTGCTCCCTCAGCTCAAGTCTGCAAGCAAGCGCGAGTTGCTCACTGCAATCA 735
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
244 -----LysAsnAsnGlnSerLeuAlaGlnAlaPheSerGluPheAlaGlnSer 259
```

```
QY 736 AAGGCAAGAAGATCACAAAGTTAGTCTCTCAATACAGTTTAGCAAAACAAGGAGATTTCG 795
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
260 LysAsnAlaThrAlaAlaLeuAlaLeuAlaTrpIleLeuAlaGlnGlyAspAspIle 279
QY 796 TCGGTGTTGGTGGGATGAGCTCTGTCTCACAGGTAGAGAAATGTTGCACAGCTTACA 855
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
280 IleProIleProGlyThrArgLysIleGluArgLeuValGluAsnAlaGlyAlaVal--- 298
QY 856 GAGCTTGAAAGTCTGGGATGATCAAGAAACTCTGTCTGAGGTTGAAGCTATTCTCGAG 915
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
299 -----AspLeuHisLeuThrAlaAlaAspLeuAlaGluIleAspAlaIleIleAla 315
QY 916 CCGTAAAGAATCTG 930
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
316 ArgTyrProAsnMet 320

RESULT 8
US-09-134-001C-3126
; Sequence 3126, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3126
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3126

Alignment Scores:
Pred. No.: 3.7e-23 Length: 314
Score: 291.50 Matches: 86
Percent Similarity: 49.84% Conservative: 72
Best Local Similarity: 27.13% Mismatches: 116
Query Match: 17.63% Indels: 43
DB: 3 Gaps: 12

US-10-606-300-12 (1-960) x US-09-134-001C-3126 (1-314)

QY 34 ACAGGGCTTAAGTTAGCGCCGCTGTTGGTGTCTCCGCTCGGAAGTGTCTTCGT 93
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
20 SerGlyIleGluIleSerGluLeuGlyLeuGlyCysMetSerLeuGlyThrAspTyr--- 38
QY 94 CCAGTCGCGGAAGATGATCGCTGCCACCGCGCGAGGCTTCCGCTCGGTATCAAC 153
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
39 -----LysLysAlaGlnProIleIleGluSerAlaIleAspAsnGlyIleThr 54
QY 154 TTCTTCGACACCTCCCGCTATTATTCGAGGAACACATGTCTGAGAAATGCTTGGTAAGGA 213
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
55 TyrPheAspThrAlaAspIleTyrAspGlnGlyValAsnGluIleValGlyLysAla 74
QY 214 CTAAAGGCTTTCGAAGTCCCTAGAAAGTCACTGCTGCTACTGCTACTAAGTGTGCTAGA--- 270
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
75 LeuLysLysTyrGln---AsnArgAspAspIleValIleGlyThrLysValGlyAsnArg 93
QY 271 ---TATAAGAGGT-----TTTGATTTTCAGTCTGAGAGAGTAAAGAGATATT 318
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
94 LeuThrAspAspGlyHisMetThrTyrAspProSerLysLysHisIleLysGluSerVal 113
QY 319 GACGAGAGCTTGAGAGGCTTCAGCTTGATTATGTTGACATATCTTCATTGCCATGACATT 378
Db |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
114 LysGlySerLeuLysArgLeuGlyLeuAsnHisLeuAspLeuTyrGlnLeuHisGly--- 132
```



```
QY 379 GAGTTCGGCTCTTGTGATCAGATTGTGAGTGAACAATTCCTGCTCTTTCAGAACTGAAA 438
Db 133 -----GlyThrIleAspProLeuAspGluThrIleSerAlaPheAspGluLeuLys 150
QY 439 CAAGAGGGGAAGACCCGGTTCATTGGTATCATCTGCTCTTCGGTTAGATATTTTCACTTAT 498
Db 151 GlnGluGlyTyrIleArgAlaTyrGlyIleSerSerIleArgProAsnValIleAspTyr 170
QY 499 GTTCTTGTGATGAGTGCCTCCAGGAGCTGCGATGTGATATTTGTCTCATCTGATTCAGGC 558
Db 171 TyrLeuLysAsn-----SerGlnIleGluThrLeuMetSer-----GlnPheAsn 185
QY 559 GTTAATGATTCGAGTTCGTGATTTACTACTTACTTGTGAGAGCAAGGTGTGGGTG 618
Db 186 LeuIleAspAsnArgProGluSerLeuIleAsnAspValHisAspLysGlnValIle 205
QY 619 ATAAGTGTCTTCCATAGCAATGGCTCTCTTACA----- 654
Db 206 LeuAlaArgGlyProValPheLysGlyLeuLeuThrSerLysSerValAspValIleAsp 225
QY 655 -----GAACAAGTCTCTGAAATGGCACCTGCTTCCCTGAGCTCAAGTCTGCA 705
Db 226 GluLysPheLysAsnGlyValLeuAspTyr-----ThrGlnAspGluLeuGlySer--- 242
QY 706 AGCAAGCCGAGTGTCTCACTGCAATCAAGGGCAGAGATCACAAAGTTAGCTCTG 765
Db 243 -----ThrIleAlaSerIleLysGluLeuGluSerAsnLeuThrAlaLeuSerPhe 259
QY 766 CAATACAGTTTAGCAACAAGAGATTTCGCGGTGTGGTGTGGATGAGCTCTGTCTCA 825
Db 260 LysTyrLeuThrSerHisAspAlaMetGlySerIleIleValGlyAlaSerSerValGlu 279
QY 826 CAGGTAGAAGAAATGTTGACAGAGTTACAGAGCTTGAAGTCTGGGGATGGATCAAGAA 885
Db 280 GlnLeuGluGluAsnValArg-----AsnTyrTyrLysGluIleSerLeuAspGln--- 296
QY 886 ACTCTGTCTGAGTGTGAGCTATTCTCGAGCTGTAAGATCTGCATGCG 936
Db 297 -----IleLysSerAlaArgAsnArgValLysAspIleGluTyr 309

RESULT 9
US-09-673-198-4
; Sequence 4, Application US/09673198
; Patent No. 6806076
; GENERAL INFORMATION:
; APPLICANT: MIYAKE, Koichiro; HASHIMOTO, Shinichi; MOTOYAMA Hiroaki;
; APPLICANT: OZAKI, Akio; SETO, Haruo; KUZAYAMA, Tomohisa; TAKAHASHI, Shunji
; TITLE OF INVENTION: A process for producing isoprenoid compounds by
; TITLE OF INVENTION: microorganisms and a method for screening compounds with
; TITLE OF INVENTION: antibiotic or weeding activity
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/673,198
; CURRENT FILING DATE: 2000-10-12
; PRIOR APPLICATION NUMBER: JP98/103101
; PRIOR FILING DATE: 1998-04-14
; PRIOR APPLICATION NUMBER: JP98/221910
; PRIOR FILING DATE: 1998-08-05
; PRIOR APPLICATION NUMBER: JP99/035739
; PRIOR FILING DATE: 1999-02-15
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-673-198-4

Alignment Scores:
Pred. No.: 4.44e-23 Length: 348
Score: 291.00 Matches: 94
Percent Similarity: 46.53% Conservative: 60
Best Local Similarity: 28.40% Mismatches: 135
Query Match: 17.60% Indels: 42
```

```
DB: 4 Gaps: 11
US-10-606-300-12 (1-960) x US-09-673-198-4 (1-348)
QY 10 ATAGAGCTTCGAGCTTTGGGGAACACAGGGCTTAAGGTTAGC-----GCC 54
Db 25 MetGlnTyrAsnProLeuGlyLysThrAspLeuArgValSerArgLeuCysLeuGlyCys 44
QY 55 GTTGGTTTGGTGTCTTCGCTCGGAGAGTGTCTTCGGTCCAGTCGCGACGAGATGATGCC 114
Db 45 MetThrPheGlyGluProAspArgGlyAsnHisAlaTrpThrLeuProGluGluSer 64
QY 115 GTCCGCCACCGTCGCGAGGCTTTCCGCTCGGTATCACTTCTTCGACACCTCCCGTAT 174
Db 65 ArgProIleIleLysArgAlaLeuGluGlyLysPheAspThrAlaAsnSer 84
QY 175 TATGGAGGAACACATGCTCTGAGAAATGCTTGGTAAGGACTAAAGGCTTTCGAAGTCCCT 234
Db 85 TyrSerAspGlySerSerGluGluIleValGlyArgAlaLeuArgAspPhe---AlaArg 103
QY 235 AGAAGTCACTACATGTTGGCTACTAAG-----TGTGTAGATATAAAGAAGGT 282
Db 104 ArgGluAspValValAlaThrLysValPheHisArgValGlyAspLeuProGluGly 123
QY 283 TTTGATTTCACTGCTGAGAGAGTAAGAAGAGTATTGACGAGAGCTTGGAGAGGCTTCAG 342
Db 124 -----LeuSerArgAlaGlnIleLeuArgSerIleAspSerLeuArgLeuGly 141
QY 343 CTTCATTTATGTTGATACATCTTTCATTTGCTGATGAGTTGGGTCTCTTGTATCAGATT 402
Db 142 MetAspTyrValAspIleLeuGlnIleHisArgTrpAspTyrAsnThr-----Pro 158
QY 403 GTGAGTGAACAATTCCTGCTCTTCAGAAATCAAGAGGGGAGAGCCCGTTCATT 462
Db 159 IledGluGluThrLeuGluAlaLeuAsnAspValValLysAlaGlyLysAlaArgTyrIle 178
QY 463 GGTATCACTGCTCTTCGCTGATATTTTCACTTATGTTCTTGTAT---CGAGTGCCTCCA 519
Db 179 GlyAlaSerSerMetHisAlaSerGlnPheAlaGlnAlaLeuGluLeuGlnHis 198
QY 520 GGGACTGCGATGTGATATTGTCTACTGTCTATACGCGC---GTTAATGATTGCGAGTTG 576
Db 199 GlyTrpAlaGlnPheValSerMetGlnAspHisTyrAsnLeuIleTyrArgGluGlu 218
QY 577 CTGAGTTTACTACCTTACTTGAAGAGCAAGGTTGGTGTGTGATAAGTGTCTTCTCATTA 636
Db 219 ArgGluMetLeuProLeuCysTyrGlnGluGlyValAlaValIleProTyrSerProLeu 238
QY 637 GCAATGGGCTCTCTTACAGAACAGGCTCTCTCTGATGGCACCTGCTTCCCTGAGCTC 696
Db 239 AlaArgGlyArgLeuThrArg-----ProTyrGlyGluThrThrAlaArgLeu 254
QY 697 AAGTCT-----GCAAGCAAAAGCCGACGTT 720
Db 255 ValSerAspGluValGlyLysAsnLeuTyrLysGluSerAspGluAsnAspAlaGlnIle 274
QY 721 GCTCAC-----TGCAAAATCAAGGGCAAGAGATCACAAAGTTAGTCTCTG 765
Db 275 AlaGluArgLeuThrGlyValSerGluGluLeuGlyAlaThrArgAlaGlnValAlaLeu 294
QY 766 CAATACAGTTTAGCAACAAGAGATTTGCTGGTGTGTGGTGGATGAGTCTCTCTCA 825
Db 295 AlaTrpLeuLeuSerLysProGlyIleAlaAlaProIleIleGlyThrSerArgGluGlu 314
QY 826 CAGGTAGAAGAAATGTTGACAGAGTTTACAGAGCTTGAAGTCTGGGATCGATCAAGAA 885
Db 315 GlnLeuAspGluLeuLeuAsnAlaVal-----AspIleThrLeuLysProGlu 330
QY 886 ACTCTGTCTGAGTGTGAAGCTATTCTCGAGCT 918
Db 331 GlnIleAlaGluLeuGluThrProTyrLysPro 341

RESULT 10
```

Qy	601	AGCAAAGCTGGGTGTGATAAAGTGGCTTCTCCATTAGCAATGGGCTCTCTTACAGAACAA	666
Db	247	ThrGluGlyIleGlySerIleAlaPheSerProLeuAlaGlyGlnLeuThrAspArg	266
Qy	661	-----	GGTCTCT
Db	267	TyrLeuAsnGlyIleProAlaAepSerArgAlaAlaSerSerSerArgPheLeuGlnPro	286
Qy	667	CCTGAATGGCACCTCGTCTCCCTCTGAG---CTCAAGTCTGCAAGCAACCGCGAGTTGCT	723
Db	287	GluGlnLeuThrProAlaArgLeuGluIleArgGlnLeuAsnArgGlnAla-----	304
Qy	724	CACTGCAAAATCAAAAGGCGACAGAGATCACAAAGTTAGTCTCTGCAATACAGTTTACCAAC	783
Db	305	-----GluAlaArgGlyGlnLysLeuSerGlnMetAlaLeuAlaTrpValLeuArgGlu	322
Qy	784	AAGGAGATTTTCGTCGGTGGTGGATGAGCTCTCTCACAGGTAGAGAAATGTT	843
Db	323	GluIysValThrSerValLeuIleGlyAlaSerLysThrAlaGlnLeuAspAsp-----	340
Qy	844	GCAGCAGTTACAGAGCTTGAAGTCTGGGATGGATCAAGAAATCTCTCTCAGGTTGAA	903
Db	341	---AlaValGlyMetLeuGlnAsnArgHisPheThrThrGluGluCysAlaAlaIleAsp	359
Qy	904	GCTATTCTC 912	
Db	360	AlaIleLeu 362	
RESULT 11			
US-09-724-623-86			
; Sequence 86, Application US/09724623			
; Patent No. 6476209			
; GENERAL INFORMATION:			
; APPLICANT: Glenn, Matthew			
; APPLICANT: Lubbers, Mark W			
; APPLICANT: Dekker, James			
; TITLE OF INVENTION: Polynucleotides, materials incorporating			
; TITLE OF INVENTION: them, and methods for using them.			
; FILE REFERENCE: 1048U1			
; CURRENT APPLICATION NUMBER: US/09/724,623			
; CURRENT FILING DATE: 2000-11-28			
; NUMBER OF SEQ ID NOS: 124			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 86			
; LENGTH: 334			
; TYPE: PRT			
; ORGANISM: Lactobacillus rhamnosus			
US-09-724-623-86			
Alignment Scores:			
Pred. No.:		1.52e-21	Length: 334
Score:		277.00	Matches: 96
Percent Similarity:		48.99%	Conservative: 74
Best Local Similarity:		27.67%	Mismatches: 107
Query Match:		16.76%	Indels: 70
DB:		4	Gaps: 18
US-10-606-300-12 (1-960) x US-09-724-623-86 (1-334)			
Qy	7	AAATAGAGCTTCGAGCTTTGGGAAACACAGGGCTTAAGTTAGCGCGTGGTTTGGT	66
Db	11	LysMetProValArgHisAlaGlyLysThrGlyLeuMetLeuProValIleSerLeuGly	30
Qy	67	GCC---TCTCGCTCGGAGTGTC-----TTCCGTCAGTCGCGAAGATGATCCCGTC	117
Db	31	LeuTrpGlnHistyryrGlyAsnLeuAspProPheGlyPro-----ArgArg	45
Qy	118	GCACCGTCGCGAGGCTTTCGGTCTCGGTATCAACTCTTTCGACACCTCCCGCTATTAT	177
Db	46	SerValIleLeuAspAlaPheAspArgGlyValPheHisPheAspValAlaAsnHistyr	65
Qy	178	GGAGGA-----ACACTGTCTGAGAAATGCTTGGTAAGGGACTA	216


```
Db 273 -----AlaHisValAlaIleAlaTrpLeuLeuSerLysSerValIleThrAlaProIle 290
QY 805 GTTGGATGAGTCTGTCTCAGGTAGAGAAATGTTGCAGCA----- 849
Db 291 IleGlyAlaThrLysProGluHisLeuSerThrAlaIleSerAlaLeuAspPheSerLeu 310
QY 850 -----GTTACAGAGCTTGAAGTCTGGGGATGGATCAAGAAACTCTGTCGAG 897
Db 311 SerAspAlaGluIleThrGluLeuGluAlaArgTyrLeuProHis-----Pro 326
QY 898 GTTGAAGCTATTCTCGAGCTCTA-----AAGAATCTGACATGGCCAAAGTGA 945
Db 327 ValAspGlyIleIleProLeuProAspThrProSerLeuThrProProSerAla 346
QY 946 ATCCATCAGAAC 957
Db 347 Ile---GlnAsn 349

RESULT 13
US-09-489-039A-12803
; Sequence 12803, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12803
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12803

Alignment Scores:
Pred. No.: 6,89e-21 Length: 328
Score: 271.00 Matches: 91
Percent Similarity: 46.32% Conservative: 60
Best Local Similarity: 27.91% Mismatches: 133
Query Match: 16.35% Indels: 42
DB: 4 Gaps: 11

US-10-606-300-12 (1-960) x US-09-489-039A-12803 (1-328)
QY 25 TTGGGGAACACAGGGCTTAAGTTAGC-----GCCGTTGGTTTTCGTGCC 69
Db 10 LeuGlyAspThrAlaLeuArgValSerArgLeuCysLeuGlyCysMetThrPheGlyGlu 29
QY 70 TCTCGCTCGAAGTGCTTCGGTCCAGTCGCGGGAAGATGATCCCGTCGCCACCGTGGCC 129
Db 30 ProAspArgGlyArgHisAlaTrpThrLeuProGluGluSerSerArgProIleGln 49
QY 130 GAGGCTTTCGGTCTCGGTATCAACTTTCGACACTCCCTCCCTATTTATGAGGAACACTG 189
Db 50 HisAlaIleGluGlyCylIleAsnPhePheAspThrAlaAsnSerTyrSerAspGlySer 69
QY 190 TCTGAGAAATCTTGGTAAAGGACTAAAGGCTTTGCAAGTCCCTAGAGTCACTACATT 249
Db 70 SerGluGluIleValGlyArgAlaLeuArgAspPhe---AlaArgAspGluValVal 88
QY 250 GTGGCTACTAAG-----TGTGGTAGATATAAAGAGGTTTGTGATTTCAGTGT 297
Db 89 ValAlaThrLysValTyrHisGlnValGlyAspLeuAlaGluGly-----LeuSerArg 106
QY 298 GAGAGAGTAAGAAGATGATTACGAGAGCTTGGAGAGCTTCAGCTTCATTATTTGAC 357
Db 107 AlaGlnIleLeuArgSerIleAspSerLeuArgLeuGlyMetAspTyrValAsp 126
QY 358 ATACTTCATTGCCATGACATTGAGTTCGGGTCTCTTGATCAGATTGTGAGTGAACAAATT 417
```

```
Db 127 LeuLeuGlnIleHisArgTrpAspTyrThrThr-----ProIleGluThrLeu 143
QY 418 CCTGCTCTTCAGAACTGAAACAAGAGGGGAAGACCCGGTTTCATTGGTATCAGTGGTCTT 477
Db 144 GluAlaLeuAspGluValValIleAlaGlyLysAlaArgTyrIleGlyAlaSerSerMet 163
QY 478 CCGTTAGATATTTTCACTTATGTTCTTCAT---CGAGTGCCTCCAGGGGACTGTCGATGTG 534
Db 164 HisAlaArgGlnPheAlaGlnAlaLeuAlaLeuGlnGlnAsnGlyTyrAlaAlaArgPhe 183
QY 535 ATATTGTCTACTGTCTTACAGGC---GTTAATGATTTCGAGCTTCTGCTGCTTACTACCT 591
Db 184 ValThrMetGlnAspHisTyrAsnLeuIleTyrArgGluGluGluAsnGluMetLeuPro 203
QY 592 TACTTGAAGACAAAGGTGGTGTGATAGTGTCTCTCCATTAGCAATGGGCTCTCTT 651
Db 204 LeuCysGlnArgAsnGlyValAlaValIleProTyrSerProLeuAlaArgGlyArgLeu 223
QY 652 ACAGAACAAAGTCTCTGTAATGGCACCTGCTTCCCTGAGCTCAAGTCTGCAAGCAA 711
Db 224 ThrArg-----ProTyrGlyGluThrThrAlaArgLeuValSerAspGluPhe 239
QY 712 GCCGAGTTGCTCACTGCAAAATCAAAG-----GGCAAG----- 744
Db 240 GlyIysSerLeuTyrSerThrSerGluGluAsnAspAlaGlnIleAlaGlyLysLeuAla 259
QY 745 -----AAGATCAACAAGTTAGTCTGTCGAATACAGATTTAGCA 780
Db 260 AspValAlaGluGluLeuAspAlaSerArgAlaGlnValAlaLeuAlaTrpLeuLeuSer 279
QY 781 AACAAAGGAGATTTCGTGGTGTGGTGGATGAGTCTGTCTCACAGGTAGAGAAAT 840
Db 280 LysProGlyValAlaAlaProIleGlyProSerArgGlnGluGlnLeuAspLeu 299
QY 841 GTTCGACGAGTTACAGAGCTTGAAGTCTGGGGATGATCAAGAACTCTGTCTCAGGTT 900
Db 300 LeuGlnAlaVal-----AspLeuThrLeuSerProGluGlnIleAspLysLeu 315
QY 901 GAAGCTATTTCGAGCCT 918
Db 316 GluAlaProTyrGlnPro 321

RESULT 14
US-09-252-991A-31535
; Sequence 31535, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31535
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31535

Alignment Scores:
Pred. No.: 4.27e-20 Length: 358
Score: 264.00 Matches: 92
Percent Similarity: 46.77% Conservative: 60
Best Local Similarity: 28.31% Mismatches: 133
Query Match: 15.97% Indels: 40
DB: 4 Gaps: 12
```

US-10-606-300-12 (1-960) x US-09-252-991A-31535 (1-358)

QY	10	ATAGAGCTTCGAGCTTTCGGGAAACACAGGGCTTAAGGTTAGCCGCGTTCGTTTGGTTGGTGC	69
Db	28	MetLysThrArgGlnLeuGlyHisAsnGlyProSerValSerAlaIleGlyLeuGlyCys	47
QY	70	TCTCGCTCGGAAGTGCTCTCGGTCCAGTCCGCCAA--GATCATGCGGTGCGCACCCGG	126
Db	48	MetGlyMetThrAspPheTyrThrThrGlyAspArgGlnGluAlaIleAlaThrLeu	67
QY	127	CGCGAGGCTTTCGGTCTCGGTATCAACTTCTTCGACACCTCCCGGTATTATGGAGGAACA	186
Db	68	HisArgAlaValGluLeuGlyLeuAsnPhePheAspThrAlaAspMetTyrGlyProHis	87
QY	187	CTGTCTGAGAAAATGCTTGTTAAGGGACTAAGGCTTTCGAAGTCCCTCAGAAGTCACTAC	246
Db	88	SerAsnGluGluLeuLeuGly-----GluAlaLeuArgGlyLysArgGluGlnVal	104
QY	247	ATTGTGGCTACTAAGTGTGGTAGATATAAAGAA-----GGTTTT	285
Db	105	PheLeuAlaSerLysPheGlyIleValArgAspProAlaAsnProArgAlaArgGlyVal	124
QY	286	GATTTCAGTCTGAGAGATGAAGAAGAGTATTTCAGAGAGCTTGAGAGGCTTCAGCTT	345
Db	125	AspGlySerProAlaTyrIleArgArgAlaIleGlySerLeuLysArgLeuGlyThr	144
QY	346	GATTATGTGCATACTTCATTCGCATGCATCAGTTCGGGTCTCTTGATCAGATT---	402
Db	145	AspArgLeuAspLeuTyrTyrGlnHisArgMetAsp-----ProGlnValPro	160
QY	403	GTGAGTGAACAATTCCTGCTCTTCAGAAACTGAACAAGAGGGGAAGACCCGGTTTCATT	462
Db	161	IleGluAspSerValGlyAlaLeuAlaAspLeuValLysAlaGlyLysIleArgHisIle	180
QY	463	GGTATCATCTGGTCTCCGTTAGATATTTTCATTATGTTCTTGATCGAGTGCCCTCCAGG	522
Db	181	GlyLeuSerGluAlaSerAlaGluThrLeuGluArgAla---HisArgValHisPro---	198
QY	523	ACTGTCGATGTGATATTCCTCATCTGTCATTCACGGCGTTAATGATTCGAGCTTGCTGGAT	582
Db	199	---IleSerAlaLeuGlnSerGluTyrSerLeuTrpThrArgAspProGluAspThrGly	217
QY	583	TTACTACCTTACTTGAAGAGCAAAAGTGTGGTGTGATAAAGTGCTTCTCCATTAGCAATG	642
Db	218	ValLeuAlaAlaCysArgArgLeuGlyIleAlaPheValProTyrSerProLeuGlyArg	237
QY	643	GGCTCTCTTACA-----GAACAAGTCTCTCTGAATGGCACCCCT	681
Db	238	GlyPheLeuThrGlyThrLeuLysArgProGluAspPheAlaAlaAspTyrArgArg	257
QY	682	GCTTCCCTC-----GAGCTCAAGTCTCGAACGAAAGCC	714
Db	258	PheSerProArgPheGlnGlyLysAsnPheAlaLysAsnLeuLysValAspLysVal	277
QY	715	GCAGTTGTCTACTGCAATCAAGAGGCAAGATCACAAGTTAGTCTCTGCAATACAGT	774
Db	278	GlyGluLeu---AlaAlaAlaLysGlyValLysProSerGlnLeuAlaLeuAlaTrpVal	296
QY	775	TTAGCAACAACAGGAGATTTCGTGGTGTGGTGGATGAGCTCTCTCAGGTAGAA	834
Db	297	LeuAlaGlnGlyAspAspLeuIleProIleProGlyThrLysGlnArgTyrThrLeuGlu	316
QY	835	GAATAATGTTGCAGCAGTTTACAGACTTGAAAGTCTGGGATGTGATCAAGAAACTTGCT	894
Db	317	GluAsnValAlaAla---ThrGluLeuArg-----LeuSerAlaAlaGluLeuAla	332
QY	895	GAGGTGTAAGCTATT	909
Db	333	GluLeuAspAlaIle	337

RESIT.T 15

RESULI 13
US-09-489-039A-9489

Sequence 9489, Application US/09489039A

```

/ Patent No. 5610836.
/
/ GENERAL INFORMATION:
/
/ APPLICANT: Gary Breston et. al
/
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO
/
/ TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOS
/
/ FILE REFERENCE: 2709.2004001
/
/ CURRENT APPLICATION NUMBER: US/09/489,039A
/
/ CURRENT FILING DATE: 2000-01-27
/
/ PRIOR APPLICATION NUMBER: US 60/117,747
/
/ PRIOR FILING DATE: 1999-01-29
/
/ NUMBER OF SEQ ID NOS: 14342
/
/ SEQ ID NO 9489
/
/ LENGTH: 347
/
/ TYPE: PRT
/
/ ORGANISM: Klebsiella pneumoniae
/
/ US-09-489-039A-9489

```

Alignment Scores:			
Pred. No.:	5.32e-19	Length:	347
Score:	254.00	Matches:	91
Percent Similarity:	45.92%	Conservative:	61
Best Local Similarity:	27.49%	Mismatches:	133
Query Match:	15.37%	Indels:	46
DB:	4	Gaps:	11
US-10-606-300-12 (1-960) x US-09-489-039A-9489 (1-347)			
QY	25	TTGGGGAAACACAGGGCTTAAGGTTAGCGCGTGTGGTTTGGTGCCTCTCCGCTCGGAAGT	84
DB	25	LeuGlyArgThrGlyLeuValSerArgLeuCysLeuGlyThrMetAsnPheGlyAsp	44
QY	85	GTCCTCGGTCCAGTCGCGAGATGATGCGGTGCCACCGTGGCGGAGGCTTTCGGTCTC	144
DB	45	Val-----ThrAspGluLeuSerArgLeuAspGluLeuGluAla	61
QY	145	GGTATCAACTTCTTCGCACACCTCCCGCTATTAT-----	177
DB	62	GlyIleAsnPheIleAspThrAlaAspValTyGlyThrGluGlnSerProAspIleGln	81
QY	178	---GGAGGAACACTGCTCTGAGAAAATCCTTGTAAGGACTAAGGCTTTCGAGTCCCT	234
DB	82	GlnGlySerGlyLeuSerGluGluIleGlyArgTrpLeuGln-----GlnGlyGly	99
QY	235	AGAAGTGACTACATTGTG---GCTACTAAG-----TGTGTAGATATAAAGAA	279
DB	100	ArgArgGluArgIleValLeuAlaThrLysValTyGlnProMetGlyProGlyProAsn	119
QY	280	GTTTTGATTTCAGTCTGAGAGTAAGAAGAGTATTGACGAGAGCTTTGGAGAGCTT	339
DB	120	AspArgArgLeuSerAlaTyrHisIleArgLysAlaCysGluAspSerLeuArgArgLeu	139
QY	340	CAGCTTGATTATGTTCATCTTCATGTCATGACATTGAGTTCCGGTCTCTTGATCAG	399
DB	140	AsnThrAspHisIleAspValTyGlnMetHisIleAspArgHisThrProTrpGlu	159
QY	400	ATTGTGAGTGAACAATTCCTGCTCTTCAGAACTGAGAGGGGAGAGCCCGTTC	459
DB	160	-----GluIleTrpGlnAlaMetGluLeuLeuValGlnGlnGlyLysValLeuTy	176
QY	460	ATTGTATCATCTGGTCTCCG---TTAGATATTTCCTTATCTTCTTCGAGTGCCT	516
DB	177	IleGlySerSerAsnPheAlaGlyTrpAspIleAlaThrAlaGlnSerValAlaThrAla	196
QY	517	CAAGGACTGTCGATGTGATTTGTCATCTGTTCATTACGGCGTTAATGATCGAGTTG	576
DB	197	ArgHisPheLeuGlyLeuValAlaGluGlnSerLeuTyAsnLeuThrAlaArgThrIle	216
QY	577	---CTGGATTTCATCTTACTTCGAGAGCAAGAGGTGTCGGTGTGATAGTCTCTCCCA	633
DB	217	GluLeuGluValIleProAlaCysArgHisPheGlyLeuGlyLeuIleProTrpSerPro	236
QY	634	TTAGCAATGGCGCTCCCTT-----ACAGAACAGAGTCTCT	666

237	LeuAlaGlyGlyLeuLeuGlyGlyValLeuLysLysMetAlaSerGlyYargArgAlaArg	256
667	CCTGAATGGCACCCCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAGCCGAGTTGCTCAC	726
257	ProAlaPheAlaArgLeuIleGluGlnTyrArgProGlnLeuGluAlaTyrGluGlyLeu	276
727	TGCAAAATCAAGGGCAAGNATCAAAAGTTAGCTCTGCAATACAGTTTACGAAACAAG	786
277	CysGluAspLeuGlyGluThrProSerAspValAlaLeuAlaTrpLeuLeuGlnAsnPro	296
787	GAGATTTTCGTCCGGTTCGTTGGGATGAGCTCTGCTCACAGGTAGAGAAATCTGCA	846
297	ValValThrAlaProLeuIleGlyProArgThrValGluGlnLeuGlnAlaLeuHis	316
847	GCAGTTACAGAGCTTGAAGCTCGGGATGGNATCAAGAACTCTCTCAGGTTGAGCT	906
317	AlaThrThr-----IleThrLeuSerAspAspThrMetSerCysLeuAspGlu	332
907	ATTCTCGAGCCTGTAAGAATCTGCATGGCCA	939
333	Ile-----TrpPro	335

Search completed: November 13, 2005, 08:32:06
Job time : 32 secs

GenCore version 5.1.6.
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 13, 2005, 08:23:17 ; Search time 99.5 Seconds
(without alignments)
8073.831 Million cell updates/sec

Title: US-10-606-300-12
Perfect score: 1653
Sequence: 1 atgacgaaatagagcttcg.....gtggaatccatcagaactaa 960

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1867879 seqs, 418409474 residues

Total number of hits satisfying chosen parameters: 3735758

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/h/US10606300/runat_07112005_094820_7339/app_query.fasta.1.1159
-DB=Published Applications AA -QMT=fastran -SUFFIX=n2p.rapp -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10606300 @CIGN 1.1.86 @runat_07112005_094820_7339
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DB TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1619	97.9	319	15	US-10-240-136A-5
2	1619	97.9	319	17	US-10-606-300-11
3	1279	77.4	335	15	US-10-424-599-262305
4	1213	73.4	316	16	US-10-437-963-113940
5	1208.5	73.1	317	16	US-10-425-115-279898
6	1199.5	72.6	335	16	US-10-767-701-44783
7	1191.5	72.1	327	16	US-10-425-115-279897
8	908	54.9	229	15	US-10-425-114-69240
9	593.5	35.9	345	20	US-11-097-143-28023
10	544.5	32.9	342	20	US-11-097-143-40479
11	420.5	25.4	294	20	US-11-097-143-71146
12	385	23.3	328	14	US-10-156-761-8943
13	368.5	22.3	328	14	US-10-156-761-14813
14	337.5	20.4	306	15	US-10-369-493-23187
15	335	20.3	945	16	US-10-437-963-171001
16	329.5	19.9	304	15	US-10-282-122A-45565
17	329	19.9	341	15	US-10-425-114-46948
18	329	19.9	341	16	US-10-425-115-357386
19	329	19.9	341	16	US-10-739-930-8425
20	327.5	19.8	325	15	US-10-369-493-4875
21	322	19.5	308	15	US-10-369-493-7635
22	322	19.5	329	15	US-10-282-122A-57831
23	316	19.1	307	15	US-10-369-493-16460
24	316	19.1	346	15	US-10-282-122A-55730
25	313	18.9	326	9	US-09-815-242-10186
26	313	18.9	326	15	US-10-369-493-862
27	313	18.9	326	15	US-10-282-122A-43244
28	310.5	18.8	302	15	US-10-282-122A-71604
29	310.5	18.8	331	14	US-10-156-761-8701
30	310	18.8	311	15	US-10-282-122A-45545
31	309.5	18.7	397	15	US-10-424-599-156741
32	308	18.6	329	15	US-10-424-599-156752
33	307.5	18.6	311	15	US-10-369-493-19379
34	306.5	18.5	332	9	US-09-815-242-11846
35	305.5	18.5	332	9	US-09-815-242-11846
36	305.5	18.5	332	15	US-10-282-122A-75573
37	302	18.3	305	15	US-10-282-122A-45821
38	301.5	18.2	339	15	US-10-369-493-20165
39	300.5	18.2	320	15	US-10-424-599-228878
40	297.5	18.0	346	15	US-10-425-114-55914
41	297.5	18.0	357	15	US-10-282-122A-57996
42	296.5	17.9	315	15	US-10-425-114-55914
43	295	17.8	344	15	US-10-369-493-12039
44	294.5	17.8	314	15	US-10-369-493-20010
45	294.5	17.8	341	16	US-10-739-930-10839

ALIGNMENTS

RESULT 1
US-10-240-136A-5
; Sequence 5, Application US/10240136A
; Publication No. US20040053235A1
; GENERAL INFORMATION:
; APPLICANT: Ascorbex Limited
; TITLE OF INVENTION: Gene Sequence
; FILE REFERENCE: JPD/P100606US-PCT
; CURRENT APPLICATION NUMBER: US/10/240,136A
; CURRENT FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: GB 0007651.3
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 5
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-240-136A-5

Alignment Scores:				
Pred. No.:	1.31e-164	Length:	319	
Score:	1619.00	Matches:	319	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	97.94%	Indels:	0	
DB:	15	Gaps:	0	
US-10-606-300-12 (1-960) x US-10-240-136A-5 (1-319)				
QY	1	ATGACGAAATAGAGCTTCGAGCTTTGGGGAACACAGGCTTAAGGTTAGCGCCGTTGGT	60	
DB	1	MetThrLysIleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGly	20	
QY	61	TTTGTGCTCTCCGCTCGGAAGTCTTCGGTCCAGTCCGCGGAAGATGATGCCGTCGCC	120	
DB	21	PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAla	40	
QY	121	ACCGTGGCGAGGCTTTCCGTCCTCGGTATCACTTCTTCGACACCTCCCGTATTATGGA	180	
DB	41	ThrValArgGluAlaPheArgLeuGlyIleAsnPheAspThrSerProTyrTyrGly	60	
QY	181	GGAAACACTGCTCGAGAAATGCTTGGTAAGGACTAAAGGCTTTGCAAGTCCCTAGAGT	240	
DB	61	GlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSer	80	
QY	241	GACTACATTGCTGCTACTAAGTGTGGTATGATATAAAGAGGTTTGTGATTTTCAGTCTGAG	300	
DB	81	AspTyrIleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGlu	100	
QY	301	AGAGTAAGAAAGATTTGACAGAGCTTGGAGAGCTTCAGCTTATGATTTGACATA	360	
DB	101	ArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle	120	
QY	361	CTTCAATTGTCATGACATTCAGTTCGGTCTCTTGATTCAGATGTCAGTGAACAATTCCT	420	
DB	121	LeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePro	140	
QY	421	GCTCTTCAGAAACTGAAACAGAGGGAAGACCCGGTTCATTTGATTCACCTGCTTCGG	480	
DB	141	AlaLeuGlnLysLeuLysGlnGlyLysThrArgPheIleGlyIleThrGlyLeuPro	160	
QY	481	TTAGATATTTTCACTTACGAGTTCAGTCCGCTCGAGTCCAGGACCTCGATGCTGATATG	540	
DB	161	LeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeu	180	
QY	541	TCATACCTGTCATTACGGCTTAATGATTCGAGTTGCTGCTGCTGCTGCTGCTGCTGCT	600	
DB	181	SerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLys	200	
QY	601	AGCAAGGTGGGTGTGATAGTCTTCTCCATTAGCAATGGGCTCCTTACAGAAACA	660	
DB	201	SerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGln	220	
QY	661	GTCCTCTGATGAGCACCTCTCCCTGAGCTCAAGTCTCGAAGCAAGCCGCTT	720	
DB	221	GlyProProGluTrpHisProAlaSerProGluLeuLysSerAlaSerLysAlaVal	240	
QY	721	GCTCACTGCAATCAAGGGCAAGAGATCAAAAGTTAGCTCTGCAATACAGTTTAGCA	780	
DB	241	AlaHisCysLysSerLysGlyLysIleThrLysLeuAlaLeuGlnTyrSerLeuAla	260	
QY	781	AACAAGGATTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	840	
DB	261	AsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGluAsn	280	
QY	841	CTTGAGGAGGATTTTCAAGCTCTGAGGATGAGTCAAGAACTCTGCTCAGGTT	900	
DB	281	ValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluVal	300	
QY	901	GAAGCTATTCTCGAGCTGTAAGAAATCTGACATGCCCAAGTGGGAATCCATCAGAAC	957	
DB	301	GluAlaIleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGlnAsn	319	

RESULT 2
 US-10-606-300-11
 ; Sequence 11, Application US/10606300
 ; Publication No. US20050019879A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Porro, Danilo
 ; APPLICANT: Sauer, Michael
 ; TITLE OF INVENTION: Ascorbic Acid Production from Yeast
 ; FILE REFERENCE: 2028.594000
 ; CURRENT APPLICATION NUMBER: US/10/606,300
 ; CURRENT FILING DATE: 2003-06-25
 ; NUMBER OF SEQ ID NOS: 26
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 11
 ; LENGTH: 319
 ; TYPE: PRT
 ; ORGANISM: Arabidopsis thaliana
 US-10-606-300-11

Alignment Scores:
 Pred. No.: 1.31e-164 Length: 319
 Score: 1619.00 Matches: 319
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 97.94% Indels: 0
 DB: 17 Gaps: 0

US-10-606-300-12 (1-960) x US-10-606-300-11 (1-319)				
QY	1	ATGACGAAATAGAGCTTCGAGCTTTGGGGAACACAGGCTTAAGGTTAGCGCCGTTGGT	60	
DB	1	MetThrLysIleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGly	20	
QY	61	TTTGTGCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCCGCGGAAGATGATGCCGTCGCC	120	
DB	21	PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAla	40	
QY	121	ACCGTGGCGAGGCTTTCCGTCCTCGGTATCAACTTCTCGACACCTCCCGTATTATGGA	180	
DB	41	ThrValArgGluAlaPheArgLeuGlyIleAsnPheAspThrSerProTyrTyrGly	60	
QY	181	GGAAACACTGCTCGAGAAATGCTTGGTAAGGACTAAAGGCTTTGCAAGTCCCTAGAGT	240	
DB	61	GlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSer	80	
QY	241	GACTACATTGCTGCTACTAAGTGTGGTATGATATAAAGAGGTTTGTGATTTTCAGTCTGAG	300	
DB	81	AspTyrIleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGlu	100	
QY	301	AGAGTAAGAAAGATTTGACAGAGCTTGGAGAGCTTCAGCTTATGATTTGACATA	360	
DB	101	ArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle	120	
QY	361	CTTCAATTGTCATGACATTCAGTTCGGTCTCTTGATTCAGATGTCAGTGAACAATTCCT	420	
DB	121	LeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePro	140	
QY	421	GCTCTTCAGAAACTGAAACAGAGGGAAGACCCGGTTCATTTGATTCACCTGCTTCGG	480	
DB	141	AlaLeuGlnLysLeuLysGlnGlyLysThrArgPheIleGlyIleThrGlyLeuPro	160	
QY	481	TTAGATATTTTCACTTACGAGTTCAGTCCGCTCGAGTCCAGGACCTCGATGCTGATATG	540	
DB	161	LeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeu	180	
QY	541	TCATACCTGTCATTACGGCTTAATGATTCGAGTTGCTGCTGCTGCTGCTGCTGCTGCT	600	
DB	181	SerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLys	200	
QY	601	AGCAAGGTGGGTGTGATAGTCTTCTCCATTAGCAATGGGCTCCTTACAGAAACA	660	
DB	201	SerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGln	220	
QY	661	GTCCTCTGATGAGCACCTCTCCCTGAGCTCAAGTCTCGAAGCAAGCCGCTT	720	
DB	221	GlyProProGluTrpHisProAlaSerProGluLeuLysSerAlaSerLysAlaVal	240	
QY	721	GCTCACTGCAATCAAGGGCAAGAGATCAAAAGTTAGCTCTGCAATACAGTTTAGCA	780	
DB	241	AlaHisCysLysSerLysGlyLysIleThrLysLeuAlaLeuGlnTyrSerLeuAla	260	
QY	781	AACAAGGATTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	840	
DB	261	AsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGluAsn	280	
QY	841	CTTGAGGAGGATTTTCAAGCTCTGAGGATGAGTCAAGAACTCTGCTCAGGTT	900	
DB	281	ValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluVal	300	
QY	901	GAAGCTATTCTCGAGCTGTAAGAAATCTGACATGCCCAAGTGGGAATCCATCAGAAC	957	
DB	301	GluAlaIleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGlnAsn	319	

QY 661 GGTCTCTCTGAATGACACCCCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAAAGCCGAGTT 720
Db 221 GlyProGluTrpHisProAlaSerProGluLeuLysSerAlaSerLysAlaAlaVal 240
QY 721 GCTCACTGCAATCAAAAGGCAAGAGATCAAAAGTTAGTCTGCAATACAGTTAGCA 780
Db 241 AlaHisCysLysSerLysGlyLysLysLysLysLysLysLysLysLysLysLysLys 260
QY 781 AACAGAGAGATTTCCTCGGTGTTGGTGGGATGAGCTCTGCTCACAGGTAGAGAAAT 840
Db 261 AsnLysGluLysSerValLeuValGlyMetSerSerValSerGlnValGluGluAsn 280
QY 841 GTTCAGCAGTTACAGAGCTTGAAGTCTGGGATGAGTCAAGAAACTCTCTCTGAGTT 900
Db 281 ValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluVal 300
QY 901 GAAGCTATTCTCGAGCTGTAAGAGTCAAGATGCGCAATGCGCAATGCAATCAGAAC 957
Db 301 GluAlaIleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGlnAsn 319
RESULT 3
US-10-424-599-262305
; Sequence 262305, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 262305
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_78884C.1.pep
US-10-424-599-262305
Alignment Scores:
Pred. No.: 5,266-128 Length: 335
Score: 1279.00 Matches: 245
Percent Similarity: 88.40% Conservative: 37
Best Local Similarity: 76.80% Mismatches: 37
Query Match: 77.37% Indels: 0
DB: 15 Gaps: 0
US-10-606-300-12 (1-960) x US-10-424-599-262305 (1-335)
QY 1 ATGACGAAATAGAGCTTCGAGTTTGGGAAACACAGGCTTAAGTTAGCCGCTGGT 60
Db 17 LeuArgGluMetGluLeuArgGluLeuGlyArgThrGlyLeuLysLeuSerThrValGly 36
QY 61 TTTGGTGGCTCTCCGCTCGAGTGTCTTCGGTCCAGTCCGCGAGATGATGCCGTCGCC 120
Db 37 PheGlyAlaSerProLeuGlyAsnValPheGlyAspValSerGluGluGlnAlaAsnAla 56
QY 121 ACCGTGCGGAGGCTTTCGGTCTCGGTATCAACTTCTCGACACCTCCCGCTATTATGGA 180
Db 57 SerValArgLeuAlaPheGlnSerGlyIleAsnPhePheAspThrSerProTyrTyrGly 76
QY 181 GGAACACTGTCTGAGAAATCTTGGTAAGGACTAAAGGCTTTGCAAGTCCCTAGAGT 240
Db 77 GlyThrLeuSerGluLysValLeuGlyLysAlaLeuLysAlaLeuGlyAlaProArgAsn 96
QY 241 GACTACATTGGCTACTAAGTCTGGTAGATATAAAGAGGTTTGTGATTTTCAAGTCTGAG 300
Db 97 SerTyrValValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGlu 116

QY 301 AGACTAAGAAAGAGTATTGACGAGAGCTTGGAGAGGCTTACGTTGATATTGTTGACATA 360
Db 117 ArgValThrArgSerIleGluGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 136
QY 361 CTTCAATGCCATGACATTCGGTCTCTTCATCAGATTCAGATTCAGTGAACAAATTCCT 420
Db 137 LeuGlnCysHisAspIleGluPheGlySerLeuAspGlnValValAsnGluThrIlePro 156
QY 421 GCTCTTCAGAACTGAAACAAAGAGGCGAAGACCCGGTTTCATTGGTATCAGTGGTCTTCG 480
Db 157 AlaLeuValLysLeuLysGluAlaGlyLysAlaArgPheIleGlyIleThrGlyLeuPro 176
QY 481 TTAGATATTTCACTTATGTTCTTTCGAGTGCCTCCAGGAGTCTCGATGATGATATG 540
Db 177 LeuGlyIlePheSerTyrValLeuAspArgValProGlyThrLeuAspValValLeu 196
QY 541 TCATACCTGCTACATTCAGGCGTTAATGATTCGACGTTGCTGGATTTACTACCTTACTTGAAG 600
Db 197 SerTyrCysHisTyrCysValAsnAspThrSerLeuGlyGluLeuValProTyrLeuLys 216
QY 601 AGCAAGGTGGGTGTGATAAGTCTTCTCATTAGCAATGGCGCTCTTACAGAACAA 660
Db 217 ThrLysGlyValGlyIleAsnAlaSerProLeuSerMetGlyLeuLeuThrGluSer 236
QY 661 GGTCTCTCTGAATGGCACCTCTTCCCTGAGTCAAGTCTGCAAGCAAAAGCCGAGTT 720
Db 237 GlyProGluTrpHisProAlaSerLeuGluLysSerAlaCysGlnAlaAla 256
QY 721 GCTCACTGCAATCAAAAGGCAAGATCAAAAGTCTGCAATGATCTGCAATACAGTTAGCA 780
Db 257 ThrHisCysLysGluLysGlyLysGlyLysAsnIleSerLysLeuAlaLeuGlnTyrSerLeu 276
QY 781 AACAGAGAGATTCGTCTCGTGTGTTGGGATGAGTCTGTCTCACAGTGAAGAAAT 840
Db 277 AsnLysGluLysThrSerValLeuValGlyMetLysSerValGluGlnValGluGluAsn 296
QY 841 GTTCAGCAGTTACAGAGCTTGAAGTCTGGGATGAGTCAAGAAACTCTCTCTGAGTT 900
Db 297 ValAlaAlaAlaArgGluLeuAlaThrSerGlyIleAspGluGluThrLeuSerGluVal 316
QY 901 GAAGCTATTCTGAGCTGTAAAGAACTGACATGCGCAAGTGGAAATCCATCAGAAC 957
Db 317 GluAlaIleLeuLysProValLysAsnGlnSerTrpProSerGlyIleGlnGlnSer 335
RESULT 4
US-10-437-963-113940
; Sequence 113940, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 113940
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_17680C.1.pep
US-10-437-963-113940
Alignment Scores:

Pred. No.: 6,48e-121 Length: 316
Score: 1213.00 Matches: 234
Percent Similarity: 83.49% Conservative: 29
Best Local Similarity: 74.29% Mismatches: 52
Query Match: 73.38% Indels: 0
DB: 16 Gaps: 0

US-10-606-300-12 (1-960) x US-10-437-963-113940 (1-316)

```
QY 10 ATAGAGCTTCAGGCTTTGGGGAACACAGGGCTTAAGGTTAGCGCGTGGTGGTGGTGGCC 69
DB 1 MetGluLeuArgGluLeuGlyAlaThrGlyLeuArgValSerProValGlyPheGlyAla 20
QY 70 TCTCGCTCGGAAGTGTCTTCGGTCCAGTCGCGGAAGATGATGCGTCCGCCACCGTGGCC 129
DB 21 SerProLeuGlyHisValPheGlyAspValProArgAspValAlaAlaValArg 40
QY 130 GAGGCTTCCGTCGCTGATCAACTTCTCGACACTCCCGTATTTAGGAGACACTG 189
DB 41 ArgAlaLeuAspLeuGlyLeuAsnPhePheAspThrSerProTyrTyrGlyThrVal 60
QY 190 TCTGAGAAATGCTTGGTAAGGACTAAAGGCTTTGCAAGTCCCTAGAAAGTCACTACAT 249
DB 61 SerGluSerValLeuGlyAspCysLeuArgAlaAlaGlyValProArgAspArgPheVal 80
QY 250 GTGGCTACTAAGTGTGTAGATATAAAGAGGTTTGTGATTTTCAGTGTGAGAGATAAGA 309
DB 81 ValAlaThrLysCysGlyArgTyrArgGluGlyPheAspPheSerAlaAlaArgValThr 100
QY 310 AAGATATTGACGAGGCTTGAGAGGCTTCAGCTTGTGATTTGATGTCATCTTCATGTC 369
DB 101 ArgSerValAspGluSerLeuAlaArgLeuGlyLeuAspTyrValAspLeuHisCys 120
QY 370 CATGACATTGCTTGGGTCCTCTTCATCAGATTGTGAGTGAACAACTTCTGCTTTCAG 429
DB 121 HisAspLeuGluPheThrAspLeuAspGlnIleValAsnGluThrIleProValLeuGln 140
QY 430 AAATGAAACAAGAGGGGAAGACCGGTTTCATGTGATACACTGCTTCGTTTCGTTAGATAT 489
DB 141 LysIleLysGluSerGlyLysAlaArgPheIleGlyIleThrGlyLeuProLeuSerIle 160
QY 490 TTCACCTATTGTTGATCGAGTGTCTCCAGGACTGTCGATGTCATATTTGTCATCTGT 549
DB 161 TyrThrTyrValLeuAspGlnValProProGlySerValAspValIleLeuSerTyrCys 180
QY 550 CATTACGGCTTAATGATTTCGACGTTGTGCTGATTTACTACCTTCTGAAGCAAGCAAGGT 609
DB 181 HisTyrGlyIleAsnAspThrAlaLeuValAspLeuLeuProTyrMetLysSerLysGly 200
QY 610 GTGGGTGTGATTAAGTGTCTTCATTAAGCAATGGGCTCTCTTACAGAACAGGTCTCTCT 669
DB 201 ValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrAspAsnGlyProPro 220
QY 670 GAATGGCACCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAGCAAGCGCAGTGTCTCAGTC 729
DB 221 GluTrpHisProAlaProLysGluLeuLysLeuAlaCysArgAlaAlaAlaAspHisCys 240
QY 730 AAATCAAGGGCAAGAAGATCAAAAGTTAGTCTGCAATACAGTTTACGAAACCAAGGAG 789
DB 241 LysLysLysGlyLysAsnIleThrLysLeuAlaMetGlnTyrSerLeuMetAsnAsnGlu 260
QY 790 ATTTTCGCTGCTGTTGGTGGAGACTCTGTCTCAGAGTGAAGAAATGTTTCAGCA 849
DB 261 IleSerThrValLeuValGlyMetAsnSerProGluGlnValGluGluAsnValAlaAla 280
QY 850 GTTACAGACTTGAAGTCTGGGATGATCAAGAACTCTCTCTGAGGTTGAAGCTATT 909
DB 281 AlaIleGluLeuSerThrSerGlyIleAspLysGluLeuHisGluValGluAlaIle 300
QY 910 CTCGAGCTTGAAGAATCTGACATGCGCCCAAGTGAATCCATCAG 954
DB 301 LeuGluProValLysAsnMetThrIrpSerSerGlyIleGluGln 315
```

RESULT 5

US-10-425-115-279898
; Sequence 279898, Application US/10425115
; Publication No. US20040214272A1

GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53222)B

; CURRENT APPLICATION NUMBER: US/10/425,115

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 369326

; SEQ ID NO 279898

; LENGTH: 317

; TYPE: PRT

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: MRT4577_186946C.1.pap

US-10-425-115-279898

Alignment Scores:

Pred. No.: 1,98e-120 Length: 317

Score: 1208.50 Matches: 236

Percent Similarity: 83.86% Conservative: 29

Best Local Similarity: 74.68% Mismatches: 50

Query Match: 73.11% Indels: 1

DB: 16 Gaps: 1

US-10-606-300-12 (1-960) x US-10-425-115-279898 (1-317)

```
QY 10 ATAGAGCTTCAGGCTTTGGGGAACACAGGGCTTAAGGTTAGCGCGTGGTGGTGGTGGCC 69
DB 1 MetGluLeuArgGluLeuGlyAlaThrGlyLeuArgValSerAlaValGlyPheGlyAla 20
QY 70 TCTCGCTCGGAAGTGTCTTCGGTCCAGTCGCGGAAGATGATGCGTCCGCCACCGTGGCC 129
DB 21 SerProLeuGlyAsnValPheGlyAspValProArgAspThrAlaAlaAlaValArg 40
QY 130 GAGGCTTCCGTCGCTGATCAACTTCTCGACACTCCCGTATTTAGGAGACACTG 189
DB 41 ArgAlaLeuAspLeuGlyIleAsnPhePheAspThrSerProTyrTyrGlyThrIle 60
QY 190 TCTGAGAAATGCTTGGTAAGGACTAAAGGCTTTGCAAGTCCCTAGAAAGTCACTACAT 249
DB 61 SerGluSerValLeuGlyAspCysLeuArgHisAlaAlaValProArgAspArgValVal 80
QY 250 GTGGCTACTAAGTGTGTAGATATAA---GAAAGTTTGGATTTTCAGTGTGAGAGACTA 306
DB 81 ValAlaThrLysCysGlyArgTyrLysAspGluGlyPheAspPheSerAlaAsnArgVal 100
QY 307 AGAAAGACTATTGACGAGAGCTTCGAGAGGCTTCAGCTTGTGATTTGTCATATCTTCA 366
DB 101 ThrArgSerIleAspGluSerLeuAlaArgLeuGlyLeuAspTyrValAspIleLeuHis 120
QY 367 TGCCATGACATTGAGTTCGGGCTCTCTTGTATCAGATTGTGAGTGAACAAATTCCTGCT 426
DB 121 AlaHisAspIleGluPheThrHisLeuGluGlnIleValAsnGluThrIleProAlaLeu 140
QY 427 CAGAACTGAAACAAGAGGGGAAGACCGGTTTCATTTGCTATCAGTGTTCCTCGCTTAG 486
DB 141 GlnLysIleLysGlyAsnGlyLysAlaArgPheIleGlyIleThrGlyLeuProLeuSer 160
QY 487 ATTTTCACTTATGTTCTTGTGATCGAGTGTCTCCAGGACTGTCCGATGTGATATTTGTC 546
DB 161 IleTyrProTyrValLeuAspArgValAlaProGlySerValAspValIleLeuSerTyr 180
QY 547 TGTCAATACCGGCTTAATGATTCGAGTGTGCTGATTTTACCTTACTTCTGAGAGCAAAA 606
DB 181 CysHisTyrGlyIleAsnAspThrSerLeuValAspLeuLeuProTyrLeuLysSerLys 200
```

QY 607 GGTGGTGGTGCATTAAGTCTTCCATTAGCAATGGCGCTCTTACAGAAAGGTCCT 666
Db GlyValGlyValIleSerAlaSerProLeuSerMetGlyLeuLeuThrAspAsnGlyPro 220
QY 667 CCTGAATGCGCCCTGCTTCCCTCAGCTCAAGTCTGCAAGCAAGCGCAGTGTCTCAC 726
Db ProGluThrHisProAlaProGluGluLeuYsSerAlaCysLysAlaAlaAlaAspHis 240
QY 727 TGCAAATCAAAAGGCGAAGATCAAAAGTGTAGCTCTGCAATACAGTTTACGAAACAAG 786
Db CysArgLysGlyLysSerIleThrLysLeuAlaMetGlnTyrSerLeuMetAsnAsn 260
QY 787 GAGATTTCGTGGTGTGGTGGATGAGCTCTGCTCAGAGTAGAAGAAAATGTTGCA 846
Db GluIleSerThrValLeuValGlyMetAsnSerLeuGluGlnValGluGluAsnValAla 280
QY 847 GCAGTTACAGAGCTTGAAGTCTGGGATGATCAAGAACTCTCTCTGAGGTTGAAGCT 906
Db AlaAlaLeuGluLeuSerThrSerGlyIleAspGluGluLeuMetArgGluValGluAla 300
QY 907 ATTCTCGAGCCTGTAAGAATCTGACATGGCCAGTGGAAATCCATCAG 954
Db IleLeuGluProValLysAsnLeuThrTrpProSerGlyIleGlnGln 316

RESULT 6

US-10-767-701-44783
; Sequence 44783, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 44783
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C59142_1.pep
US-10-767-701-44783

Alignment Scores:

Pred. No.: 1.89e-119 Length: 335
Score: 1199.50 Matches: 233
Percent Similarity: 83.86% Conservative: 32
Best Local Similarity: 73.73% Mismatches: 50
Query Match: 72.57% Indels: 1
DB: 16 Gaps: 1

US-10-606-300-12 (1-960) x US-10-767-701-44783 (1-335)

QY 10 ATAGAGCTTCGAGCTTCGGGGAACACAGCGCTTAAGTTAGCGCGTGGTTTGGTGCC 69
Db MetGluLeuArgGluLeuGlyThrGlyLeuArgValSerAlaValGlyPheGlyAla 38
QY 70 TCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCCGGAAGATGATGCCGTCGCCACCGTCGCC 129
Db SerProLeuGlyAsnValPheGlyAspValProArgAspThrAlaArgAlaAlaValArg 58
QY 130 GAGGCTTTCGCTTCGGTATCACTCTTCGACACCTCCCGTATTATGGAGGACACTG 189
Db ArgAlaLeuAspLeuGlyIleAsnPheAspThrSerProTyrTyrGlyThrVal 78
QY 190 TCTGAGAAAATCTCTGGTAAAGGACTAAAGGCTTTGCAAGTCCCTAGAAAGTGAATACAT 249
Db SerGluSerValLeuGlyAspCysLeuArgHisAlaAlaValProArgAspArgValVal 98
QY 250 GTGGCTACTAAGTGTGTAGATATAAA---GAAGGTTTGTGATTTTCAGTGTCTGAGAGATA 306

Db ValAlaThrLysCysGlyArgTyrLysAspGluGlyPheAspPheSerAlaAspArgVal 118
QY 307 AGAAAGAGTATTGACGAGAGCTTCGAGAGGCTTCAGCTTGATTTGTTGACATCTTCAT 366
Db ThrArgSerIleAspGluSerLeuAlaArgLeuGlyLeuAspTyrValAspIleLeuHis 138
QY 367 TGCATGACATGATGAGTTCGGGTCTCTTGATCAGATTGTGAGTGAACAATTCCTGCTCTT 426
Db AlaHisAspIleGluPheThrHisLeuAspGlnIleValAsnGluThrIleProAlaLeu 158
QY 427 CAGAACTGAAACAGAGGGAAGACCGCTTCATTCGATCAGCTGCTTCCTTCCTAGAT 486
Db GlnLysIleLysGluSerGlyLysAlaArgPheIleGlyIleThrGlyLeuProLeuSer 178
QY 487 ATTTTCTACTTATGTTCTTTCGATGAGTCTCCAGGACTGTCGATGTGATATTGTCTATC 546
Db IleTyrProTyrValLeuAspArgValAlaProGlySerValAspLeuLeuSerTyr 198
QY 547 TGTCAATTACGCGCTTAATGATTCGACGTTGCTGGATTTACTACCTTACTTGAAGACAAA 606
Db CysHisTyrGlyIleAsnAspThrSerLeuValAspLeuLeuProTyrLeuLysSerLys 218
QY 607 GGTGGTGGTGTGATAAGTCTTCTCCATTAGCAATGGCGCTCTTACAGAAACAGGTCCT 666
Db GlyValGlyValIleSerAlaSerProLeuSerMetGlyLeuLeuThrAspAsnGlyPro 238
QY 667 CTCAATGGCACTGCTTCCCTGAGCTCAAGTCTGCAAGCAAGCGCAGTGTCTCAC 726
Db ProGluThrHisProAlaProGluGluLeuYsSerAlaCysArgAlaAlaAlaAspHis 258
QY 727 TGCAAATCAAAAGGCGAAGATCAAAAGTGTAGCTCTGCAATACAGTTTACGAAACAAG 786
Db CysArgLysGlyArgSerIleThrLysLeuAlaMetGlnTyrSerLeuMetAsnAsn 278
QY 787 GAGATTTCGTGGTGTGGTGGATGAGCTCTCTCAGAGTAGAAGAAAATGTTGCA 846
Db GluIleSerThrIleLeuValGlyMetAsnSerSerGluGlnValGluGluAsnValAla 298
QY 847 GCAGTTACAGAGCTTGAAGTCTGGGATGATCAAGAACTCTCTCTGAGGTTGAAGCT 906
Db AlaAlaLeuGluLeuSerThrSerGlyIleAspGluGluLeuArgGluValGluAla 318
QY 907 ATTCTCGAGCCTGTAAGAATCTGACATGGCCAGTGGAAATCCATCAG 954
Db IleLeuGluProValLysAsnLeuThrTrpProSerGlyIleGlnGln 334

RESULT 7

US-10-425-115-279897
; Sequence 279897, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 279897
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_186845C.1.pep
US-10-425-115-279897

Alignment Scores:

Pred. No.: 1.36e-118 Length: 327
Score: 1191.50 Matches: 236


```
/ PRIOR APPLICATION NUMBER: 60/175,693
/ PRIOR FILING DATE: 2000-01-12
/ PRIOR APPLICATION NUMBER: 60/184,831
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: 60/191,637
/ PRIOR FILING DATE: 2000-03-23
/ NUMBER OF SEQ ID NOS: 43008
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 40479
/ LENGTH: 342
/ TYPE: PRT
/ ORGANISM: DROSOPHILA
US-11-097-143-40479

Alignment Scores:
Pred. No.: 5,95e-49 Length: 342
Score: 544.50 Matches: 117
Percent Similarity: 58.75% Conservatives: 71
Best Local Similarity: 36.58% Mismatches: 117
Query Match: 32.94% Indels: 15
DB: 20 Gaps: 7

US-10-606-300-12 (1-960) x US-11-097-143-40479 (1-342)
QY 7 AAATAGAGCTTCGAGCTTTGGGGAACACAGGCTTAAGTTAGCGCGTGTGGTTTGGT 66
Db 21 ArgMetGluTyrArgGlnLeuGlySerThrGlyLeuArgValSerLysAlaLeuGly 40
QY 67 GCCTCTCCGCTCGGAAGTCTCTTCGGT---CCAGTCGCGAAGATGATCGCGTCCGCCACC 123
Db 41 GlyAlaThrLeuSerLysLeuPheSerAspPheAspArgGluGluGlyIleLeuThr 60
QY 124 GTGCGGAGGCTTCCGCTCTCGGTATCAACTTCTTCGACACCTCCCGTATTATGAGGA 183
Db 61 ValGlnAlaIleArgSerGlyIleAsnTyrIleAspThrAlaProPheThrGlyGln 80
QY 184 ACACGTCTGAGAAATGCTTGGTAAGGACTAAAGGCTTCCAAAGTCTCCAGAGTAC 243
Db 81 GlyLysSerGluLeuLeuGlyGlnAlaLeuLys-----AspValProArgGluAla 98
QY 244 TACATTGTGGCTACTAAGTGTGTAGATATAA-----GAAGTTTTCATTTC 291
Db 99 TyrTyrIleAlaThrLysValAlaArgTyrGluLeuAspProAsnAsnMetPheAspTyr 118
QY 292 AGTGTCTGAGAGTAAGAAGATATTGACGAGCTTGGAGAGCTTCAGAGCTTCAGTTGATTAT 351
Db 119 ThrAlaAlaLysAlaArgLysSerValLysArgSerLeuGluLeuGlnLeuAspArg 138
QY 352 GTTGACATATTCATTGCCATGACATTCAGTTGCGGG---TCTCTGATCAGATTGTGAGT 408
Db 139 ValAspValLeuGlnValHisAspValAlaProSerLeuAspMetValLeuAsn 158
QY 409 GAAACAATTCTCTCTCAGAACTGAAACAAGAGGGGAGACCCGGTTTCATTGTTATC 468
Db 159 GluThrIleProValLeuGluGluTyrValGlnAlaGlyLysAlaArgPheIleGlyVal 178
QY 469 ACTGGTCTTCCTGATATTTTCATTATGTTTCTTGTATCGAGTGCCTCCAGGGACTGTC 528
Db 179 ThrAlaTyrAspValAspValLeuLysGluCysAlaGluArg--GlyLysGlyArgIle 197
QY 529 GATGTGATATTGTCATATGTCATTCACGCGGTTATGATTCACGTTGCTGATTTACTA 588
Db 198 GlnValValLeuAsnTyrAlaArgTyrThrLeuLeuAspAsnThrLeuLeuArgHisMet 217
QY 589 CCTTACTTCAAGAGCAAGAGTGTGGGTGTGATAAGTGTCTTCCTATTAGCAATGGCGCTC 648
Db 218 LysAlaPheGlnGluMetGlyValGlyValValCysAlaAlaAlaHisSerLeuGlyLeu 237
QY 649 CTTACAGAACAAAGTCTCTCTGAATGGCACCTCTGCTTCCCTGAGCTCAAGTCTCAAGC 708
Db 238 LeuSerAsnAlaGlyProGlnSerTrpHisProGlySerProGluLeuLeuAlaValGly 257
QY 709 AAAGCGCGATGCTCACTGCAAAATCAAGGGCAAGAGATCACAAGTTAGTCTGCNA 768

Db 258 LysArgGlyAlaGluIleCysGlnLysArgAsnValGluLeuGlyLysLeuAlaMetTyr 277
QY 769 TACAGTTTAGCAAAACACAGGAGATTTCGTCGCTGTTGGTTGGGATGAGCTCTGTCTCACAG 828
Db 278 TyrThrMetGlnLeuAspGlyAlaThrPheLeuIleGlyIleProAsnArgLysLeu 297
QY 829 GTAGAAGAAATGTTGACAGCTTACAGAGCTTGAAAGTCTGGGATGATGATCAAGAACT 888
Db 298 LeuArgIleAsnLeuAspAlaIlePheAsp-----GlyLeuThrSerHisGlu 313
QY 889 CTGCTGAGGTTGAAGCTATTCTGAGCCTGTA-----AAGAACTCTCACATGGCCAACT 942
Db 314 GlnGluValLeuGlnTyrLeuArgGluAsnValPheThrLysSerTyrSerTrpGlySer 333

RESULT 11
US-11-097-143-7146
/ Sequence 7146, Application US/11097143
/ Publication No. US20050208558A1
/ GENERAL INFORMATION:
/ APPLICANT: Venter, J. Craig
/ APPLICANT: et al.
/ TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
/ TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
/ FILE REFERENCE: CL000728
/ CURRENT APPLICATION NUMBER: US/11/097,143
/ CURRENT FILING DATE: 2005-04-04
/ PRIOR APPLICATION NUMBER: 60/157,832
/ PRIOR FILING DATE: 1999-10-05
/ PRIOR APPLICATION NUMBER: 60/160,191
/ PRIOR FILING DATE: 1999-10-19
/ PRIOR APPLICATION NUMBER: 60/161,932
/ PRIOR FILING DATE: 1999-10-28
/ PRIOR APPLICATION NUMBER: 60/164,769
/ PRIOR FILING DATE: 1999-11-12
/ PRIOR APPLICATION NUMBER: 60/173,383
/ PRIOR FILING DATE: 1999-12-28
/ PRIOR APPLICATION NUMBER: 60/175,693
/ PRIOR FILING DATE: 2000-01-12
/ PRIOR APPLICATION NUMBER: 60/184,831
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: 60/191,637
/ PRIOR FILING DATE: 2000-03-23
/ NUMBER OF SEQ ID NOS: 43008
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 7146
/ LENGTH: 294
/ TYPE: PRT
/ ORGANISM: DROSOPHILA
US-11-097-143-7146

Alignment Scores:
Pred. No.: 1.22e-35 Length: 294
Score: 420.50 Matches: 102
Percent Similarity: 52.48% Conservatives: 57
Best Local Similarity: 33.66% Mismatches: 101
Query Match: 25.44% Indels: 43
DB: 20 Gaps: 7

US-10-606-300-12 (1-960) x US-11-097-143-7146 (1-294)
QY 7 AAATAGAGCTTCGAGCTTTGGGGAACACAGGCTTAAGTTAGCGCGTGTGGTTTGGT 66
Db 21 ArgMetGluTyrArgGlnLeuGlySerThrGlyLeuHisValSerLysLeuAlaIleGly 40
QY 67 GCCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGAAGATGATGTCGCTCCGCCCGTG 126
Db 41 GlySerProLeuCysAsnLeu----- 47
QY 127 CGCAGGCTTTCCCGTCTCGGTATCAACTTCTTCGACACCTCCCGTATTATGAGGAACA 186
Db 48 -----PhePheAspAspTyrAspArgGluGluGlyIle 58
```



```
QY 766 CAATACAGTTTACCAACAGAGATTTCCTCGTGGTGGTGGATGAGCTCTGCTCA 825
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 275 ArgPheProLeuArgAspProSerValAlaCysValValAlaGlyPheArgThrSerGlu 294
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 826 CAGGTAGAGAAATGTTGCA---GCAGTTACAGAGCTTTGAAAGTCTCGGGATGATCAA 882
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 295 GluValValSerAlaAlaArgTrpAlaThrThrAspLeuThrGly----- 309
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 883 GAAACTGTCTGAGGTTGAGCTATTCTCGAGCCTGTAAAGAACTCTGACATGGCCAAAGT 942
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 310 -----GluAlaTrpLeuAspLeuAspThrAlaAlaThrTrpAlaAla 323
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 943 GGAATC 948
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 324 GluIle 325

RESULT 13
; Sequence 14813, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: HORIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14813
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14813

Alignment Scores:
Pred. No.: 5,096-30 Length: 328
Score: 368.50 Matches: 101
Percent Similarity: 50.68% Conservative: 49
Best Local Similarity: 34.12% Mismatches: 115
Query Match: 22.23% Indels: 31
DB: 14 Gaps: 7

US-10-606-300-12 (1-960) x US-10-156-761-14813 (1-328)
QY 25 TTGGGAACACAGGCTTAAGGTAGCGCGTGTGTTGGTCTCTCGCTCGGAAGT 84
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 LeuGlyArgSerGlyValGluValSerGluLeuSerPheGlyAlaAlaGlyIleGlyAsn 23
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 85 GTCTTCGTCCAGCCGCAAGATGATCGCGTCCGCCACCGCGCGAGGCTTTCGCTC 144
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 24 LeuPheThrAlaValThrAspGluGlnAlaSerGlnAlaValHisAlaAlaTrpThrSer 43
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 145 GGTATCAACTTCTCGACACCTCCCGCTATATATGAGGAACACTGTCTGAGAAATGCTT 204
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 44 GlyIleArgTyPheAspThrAlaProHisIleTyGlyLeuGlyLeuSerGluArgLeu 63
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 205 GGTAAAGGACTAAAGGCTTTGCAAGTCCCTAGAGTGACTACATTGTGCTACTAAGTGT 264
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 64 GlyAlaAlaLeuArg-----GluHisProArgThrAlaTyThrValSerThrIysVal 81
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 265 GGTAGATATAAGAA-----GGTTTT----- 285
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 82 GlyArgArgLeuGluProThrAspAlaGlyGlyAspAspLeuAlaAspGlyPheAlaVal 101
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY 286 -----GATTTCAGTGTGAGAGAGTAAGAAAGAGATTTCAC 321
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 102 ProAlaThrHisHisArgValTrpAspPheSerAlaAspGlyValArgArgAlaLeuThr 121
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 322 GAGAGCTTGGAGAGGCTTACGCTTGTATTTGACATATCTTATTTGCCATTCGACATTGAG 381
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 AlaSerLeuGluArgLeuGlyLeuAspArgValValValValValLeuHisAspProAsp 141
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 382 TTCGGGCTCTTGTGATTCAGATTGTGAGTGAACAATCTCTCTCTTCAGAAATCGAAACAA 441
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 142 ---AspHisAlaGluGlnAlaPheArgGluGlyCysProAlaLeuGluGlyLeuArgSer 160
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 442 GAGGGAAGACCCGGTTTCATTGGTATCACCTGCTTCCGTTAGATATTTTTCATTCTTGT 501
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 161 GluGly-----ValValGlyAlaIleGlyAlaGlyMetAsnGlnThrAlaMetLeu 177
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 502 CTTGATCGAGTCCCTCCAGGACTGTGATGTGATATTGTCTATCTGTCTATTACGCGCTT 561
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 178 ThrArgPheValArgGluThrAspValValValValValValValValValValValVal 197
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 562 AATGATTCG---ACGTTCTGCTGATTTACTACTTCTTCTTGAAGAGCAAGGTTGGGTGTG 618
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 198 LeuAspGlnArgAlaLeuAlaAspLeuLeuProAlaAlaAlaAspArgGlyThrSerVal 217
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 619 ATAAGTGCTTCTCCATTAGCAATGGCTCTTACAGAACAAAGGTCCTCCT- 669
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 218 ValIleGlyGlyAlaPheAsnSerGlyLeuLeuAlaAspProLysProGlyAlaThrTyr 237
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 670 GAATGGCACCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAAAGCGCAGTTGCTCAGTCG 729
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 AsnTyAlaAlaAlaProProGluLeuLeuAspArgAlaLeuArgLeuLysAlaValAla 257
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 730 AATCAAGGCAAGAGATCACAAGTTAGTCTCTGCAATACAGTTTACGAAACAAAGGAG 789
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 258 GluArgHisGlyThrThrLeuArgGlyAlaAlaLeuAlaPheCysAlaAlaHisProAla 277
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 790 ATTTTCGTCGGTGGTGGATGAGCTCTCTCACAGGTAGAGAA 837
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 278 ValAlaSerValLeuValGlyAlaArgSerProTyArgGluValAlaArgAsp 293
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
US-10-369-493-23187
; Sequence 23187, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 23187
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-10-369-493-23187

Alignment Scores:
Pred. No.: 1,07e-26 Length: 306
Score: 337.50 Matches: 100
Percent Similarity: 51.41% Conservative: 64
Best Local Similarity: 31.35% Mismatches: 108
Query Match: 20.42% Indels: 47
DB: 15 Gaps: 12

US-10-606-300-12 (1-960) x US-10-369-493-23187 (1-306)
```


QY 19 CGAGCTTTGGGAACACAGGGCTTAAGGTTAGCGCGTTGGTGGTCTTCGCGCTC 78
Db 4 ArgLysLeuGlyThrSerAspLeuAspIleSerGluValGlyLeuGlyCysMetSerLeu 23
QY 79 GGAAGTGTCTCGTCCAGTCCGCAAGATGATGCGTCCGACCGTGGCGAGGCTTTC 138
Db 24 GlyThr-----GluYsAsnLysAlaLeuSerIleLeuAspGluAlaIle 38
QY 139 CGTCTCGGTATCACTTCTTCACACACCTCCCGTATTATGGAGGAACACTGTCTGAGAA 198
Db 39 GluLeuGlyIleAsnTyrLeuAspThrAlaAspLeuTyrAspArgGlyArgAsnGluGlu 58
QY 199 ATGCTTGGTAGGAGCTAAGGCTTTCAGAGTCCCTAGAGTGACTACATTTGGCTACT 258
Db 59 IleValGly-----AspAlaIleGlnAsnArgArgHisAspIleLeuAlaThr 75
QY 259 AGTGTGCT---AGATATAAGAAAGT-----TTTGATTTTCAGTCTGAG 300
Db 76 LysAlaGlyAsnArgTrpAspAspGlySerGluGlyTrpTyrTrpAspProSerLysAla 95
QY 301 AGAGTAAGAAAGATGATTGACAGAGCTTGGAGAGCTTCAGCTTGATTTGTTGACATA 360
Db 96 TyrIleLysGluAlaValLysSerLeuThrArgLeuLysThrAspTyrIleAspLeu 115
QY 361 CTTCAATGCCATGACATTCAGTTCGGGTCTCTTGATCAGATTGTGAGTGAACAAATTCCT 420
Db 116 TyrGlnLeuHisGly-----GlyThrIleGluAspAsnIleAspGluThrIleGlu 132
QY 421 GCTCTTCAGAACTGAACACAGAGGGGAGACCGGTTTCATTTGATCAGTGTCT--- 477
Db 133 AlaPheGluGluLeuLysGlnGluGlyValIleArgTyrTyrGlyIleSerSerIleArg 152
QY 478 CGGTTAGATATTTTCACCTATGTTCTTGTGAGTGCCTCCAGGAGCTGTCGATGTGATA 537
Db 153 ProAsnValIleLysGluTyrValLysLysSer-----AsnIleVal 166
QY 538 TTGTGCATCTGTCAATAGCGGCTTAATGATTCGAGCTTGTGGATTTACTACTTACTTG 597
Db 167 SerIleMetMetGlnPheSerLeuPheAspArgArgProGluGluTrpLeuProLeuLeu 186
QY 598 AAGCAAAAGTGGGTGTGATAGTGTCTTCATAGCAATGGGCTCTCTACAGAA 657
Db 187 GluGluHisGlnIleSerValValAlaArgGlyProValAlaLysGlyLeuLeuThrGlu 206
QY 658 CAAGGTCTCTCGAATGGCACCTCTCTCC----- 687
Db 207 Lys-----ProLeuAspGlnAlaSerGluSerMetLysGlnAsnGlyTyrLeuSerTyrSer 225
QY 688 ---CCTGAGCTCAAGTCCGAAGCAAGGCC-----GCAGTTGTCTCACTGCAAAATCAAAG 738
Db 226 PheGluGluLeuThrAsnAlaArgLysAlaMetGluGluValAlaProAspLeuSer--- 244
QY 739 GCGAAGAAGATCAAAAGTTAGCTCTGCAATACAGTTTACAGAAACAGAGAGATTCGTG 798
Db 245 -----MetThrGluLysSerLeuGlnTyrLeuLeuAlaGlnProAlaValAlaSer 261
QY 799 GTGTTGTTGGATGAGTCTGTCTCAGAGTGAAGAAATGTTGACGCA----- 849
Db 262 ValIleThrGlyAlaSerLysIleGluGlnLeuArgGluAsnIleGlnAlaAlaAsnAla 281
QY 850 -----GTTACAGACTTCAAGTCTGGGGATGGATCAAGAAACTCTCTCTGAGGTT 900
Db 282 ArgArgLeuThrGluGluGluIleLysAlaLeuGlnSerHisThrLysGlnAspIle 300

RESULT 15

US-10-437-963-171001
; Sequence 171001, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 171001
; LENGTH: 945
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_69273C.1.pbp
US-10-437-963-171001

Alignment Scores:

Pred. No.: 3,52e-26 Length: 945
Score: 335.00 Matches: 107
Percent Similarity: 47.06% Conservative: 53
Best Local Similarity: 31.47% Mismatches: 118
Query Match: 20.27% Indels: 62
DB: 16 Gaps: 14

US-10-606-300-12 (1-960) x US-10-437-963-171001 (1-945)

QY 25 TTGGGGAAACACAGGGCTTAAGGTTAGCGCGCTTGGTGGTCTTCGCGTCCGGAAGT 84
Db 546 LeuGlyThrGlnGlyLeuGluValSerLysLeuGlyPheGlyCysMetGlyLeuThrGly 565
QY 85 GTCTTC---GGTCAGTCGCCGAAGATGATCCGTCGCCACCGTCGCGAGGCTTTCGCT 141
Db 566 ValTyrAsnAlaProValAlaGluAspGlyValAlaValAlaValArgAlaPheGlu 585
QY 142 CTCGGTATCACTTCTTCGACACCTCCCGCTATTATGGAGGAACACTCTCTGAGAAATG 201
Db 586 AlaGlyValThrPhePheAspThrSerAspAlaTyrGlyProHisThrAsnGluValLeu 605
QY 202 CTTGGTAGGAGCAATAAGGCTTTCAGATCCTCCTAGAGTACTACATTTGGTCTACTAAG 261
Db 606 LeuGlyLysAlaLeuLys-----GlnLeuProArgGluLysValGlnValAlaThrLys 623
QY 262 TGTGTAGATATATAAGAGGTTTGTATTTCAGTCT-----GAG 300
Db 624 PheGly-----IleAlaGlyPheAspAlaAsnGlyMetLeuValLysGlyThrProAsp 641
QY 301 AGAGTAAGAAAGATGATTTCGACGAGCTTGGAGAGGCTTCAGCTTGTATTGTTGACATA 360
Db 642 TyrValArgAlaCysCysGluAlaSerLeuGluArgLeuAlaValAspTyrIleAspLeu 661
QY 361 CTTCAATGCCATGACATTCAGTTCGGGTCTCTTGATCAGATTGTG-----AGTGAACA 414
Db 662 TyrTyrGlnHisArgIle-----AspGlnSerValProIleGluGluThr 676
QY 415 ATTCTGCTCTTCAGAACTGAAACAGAGGGGAGACCCGTTTCATTGTTGATCACTGCT 474
Db 677 MetGlyGluLeuLysLysLeuValGluGluGlyLysValLysPheValGlyLeuSerGlu 696
QY 475 CTTCCGTTAGATATTTTCACTTATGTTCTTGATCGAGTCCCTCCAGGAGCTGTGATGTG 534
Db 697 AlaSerAlaAspThrIleArgArgAlaHisAlaValTyrProIleThrAlaValGlnMet 716
QY 535 ATATTGTCATCTGTCATTACGGGCTTAATGATTCGAGCTTGGTGGATTTACTACTTAC 594
Db 717 GluTrpSerLeuTrpThrArgAspIleGlu-----GluIleIleProLeu 732
QY 595 TTGAAG-----ACGAAAGGTGTGCTGATAGTCTCTCCATTAGCAATGGGCTC 648
Db 733 CysArgPheMetGluLeuGlyIleGlyValProTyrSerProIleGlyArgGlyPhe 752

```
QY 649 CTTACAGAACAAAGGTCCTCCTCAATGGCACCCCTGCTTCCCTGAGCTCAAGTCTCAAGC 708
Db   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::
753 PheAlaGlyArgAlaAlaValGlnSerIleProSerGluSerTrpLeu----- 768
QY 709 AAAGCGCAGTTGCTCACTGCAATCAAAGGCAAG----- 744
Db   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::
769 -----ThrArgHisProArgTyrAsnGlyGluAsnLeuGluLysAsnLysValPhe 785
QY 745 -----AAGATCACAAAGTTAGTCTGCAATAC-----AGTTTAGCAAAACAAGGAGATT 792
Db   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::
786 TyrThrArgIleGluGluLeuAlaThrLysTyrGlyCysSerProAlaGlnLeuAlaLeu 805
QY 793 TCGTCGGTGTG-----GTTGGGATGAGCTCTGCTCTCA 825
Db   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::
806 SerTrpValLeuHisGlnGlyAspValValProIleProGlyThrThrLysValLys 825
QY 826 CAGGTAGAGAAATGTTGCAGCAGTTACAGAGCTTGAAGTCTGGGATGGATCAAGAA 885
Db   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::
826 AsnLeuAspAspAsnIleGlyAlaVal-----LysValLysLeuSerLysGlu 841
QY 886 ACTCTGTCTGAGTTGAAGCTATTCTC-----GAGCCTGTAAGAATCTG 930
Db   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::
842 AspLeuLysGluIleSerAlaAlaValProAlaGlyGluGluProArgGluGlnLeu 861
```

Search completed: November 13, 2005, 08:35:31
Job time : 108.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 13, 2005, 08:03:46 ; Search time 25.5 Seconds
(without alignments)
7244.556 Million cell updates/sec

Title: US-10-606-300-12

Perfect score: 1653

Sequence: 1 atgcgaaatagagcttcg.....gtggaatccatcagaactaa 960

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO-spool_h/US10606300/runat_07112005_094819_7311/app_query.fasta_1.1159
-DB=PIR -QFMT=fastan -SUFFIX=en2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10606300 @CGN 1 1 27 @runat_07112005_094819_7311 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR 79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1619	97.9	319	2	T04984
2	503	30.4	439	2	T28841
3	366	22.1	329	2	JC2405
4	362.5	21.9	297	2	C83776
5	358	21.7	335	2	S82890
6	346.5	21.0	329	2	Af2990
7	346.5	21.0	329	2	C98293
8	337.5	20.4	306	2	H69966
9	317	19.2	351	2	A82979
10	317	19.2	351	2	H98303
11	314	19.0	326	2	H90938
12	314	19.0	326	2	D85787
13	313	18.9	326	2	C64937
14	305.5	18.5	323	2	H83427

15	305.5	18.5	332	2	AB0808	probable ion-chann
16	305	18.5	319	2	H72391	k+ channel, beta s
17	304.5	18.4	338	2	G95357	probable [imported
18	301.5	18.2	332	2	D98252	general stress pro
19	301.5	18.2	332	2	AH3033	aldo-keto reductas
20	298.5	18.1	353	2	AH3156	oxidoreductase moc
21	298.5	18.1	353	2	A98131	mocA protein [impo
22	292	17.7	310	2	D69646	myo-inositol catab
23	291	17.6	324	2	A90688	probable NAD(P)H-d
24	291	17.6	324	2	E85538	probable NAD(P)H-d
25	291	17.6	348	2	C64771	probable oxidoredu
26	290	17.5	319	2	T35337	probable ion chann
27	289	17.5	369	2	C97390	aldo/keto reductas
28	287.5	17.4	346	2	E91114	probable reductase
29	287	17.4	333	2	AD2608	aldo-keto reductas
30	286.5	17.3	327	2	E75296	aldo/keto reductas
31	286.5	17.3	338	2	AI3183	aldo/keto reductas
32	284	17.2	301	2	B86986	probable oxidoredu
33	284	17.2	324	2	AE0554	probable oxidoredu
34	283	17.1	310	2	B69755	ion channel homolo
35	281.5	17.0	346	2	E85959	probable reductase
36	277.5	16.8	346	2	G65086	hypothetical prote
37	276.5	16.7	329	2	E82644	sugar-phosphate de
38	276.5	16.7	362	2	D82644	sugar-phosphate de
39	276	16.7	353	2	F95403	probable oxidoredu
40	274.5	16.6	331	2	B95902	probable aldoketo
41	274.5	16.6	354	2	D90187	oxidoreductase [im
42	271.5	16.4	379	2	E72284	oxidoreductase, al
43	270.5	16.4	340	2	B96632	hypothetical prote
44	270	16.3	329	2	AI0341	probable aldo/keto
45	268.5	16.2	315	2	A75289	probable potassium

ALIGNMENTS

RESULT 1

T04984

hypothetical protein T16L1.160 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004

C:Accession: T04984

R:Bevan, M.; Obermaier, B.; Deutschenbaun, S.; Piravandi, E.; Hoheisel, J.; Mewes, H.W.;

submitted to the Protein Sequence Database, November 1998

A:Reference number: Z15393

A:Accession: T04984

A:Molecule type: DNA

A:Residues: 1-319 <BEV>

A:Cross-references: UNIPROT:O81884; EMBL:AL031394

A:Experimental source: cultivar Columbia; BAC clone T16L1

C:Genetics:

A:Map position: 4

A:Introns: 57/2; 133/3; 233/3; 276/3

A>Note: T16L1.160

C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Alignment Scores:			
Pred No.:	1.38e-123	Length:	319
Score:	1619.00	Matches:	319
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	97.94%	Indels:	0
DB:	2	Gaps:	0

US-10-606-300-12 (1-960) x T04984 (1-319)

Qy	1	ATGACGAAATAGAGCTTCGAGCTTTGGGAAACACAGGCTTAAGGTTAGCGCGCTTGGT	60
Db	1	MetThrIysileGluLeuArgAlaLeuGlyAenThrGlyLeuLysValserAlaValGly	20
Qy	61	TTTTCGTCCTCCGCTCGGAGTGTCTTCGTCAGTCGCGAAGATGATGCGCTCGCC	120
Db	21	PheGlyAlaserProLeuGlySerValPheGlyProValAlaGluAspAlaValAla	40

QY 121 ACCGTGCGGAGGCTTTCCGTCTCGGTATCAACTTCTTCGACACCTCCCGGTATTTATGGA 180
Db 41 ThrValArgGluAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyrTyrGly 60
QY 181 GGACACACTGCTGAGAAATGCTTGGTAAAGGACTTAAAGGCTTTCGAAGTCCCTAGAGT 240
Db 61 GlyThrLeuSerGluLeuMetLeuGlyLysGlyLysAlaLeuGlnValProArgSer 80
QY 241 GACTACATTGCTGCTACTAAGTGTGGTATAGATATAAAGAGGTTTGTGATTTGAGTCTCGAG 300
Db 81 AspTyrIleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGlu 100
QY 301 AGAGTAAGAAAGATGATTGACGAGACTTGGAGAGGCTTCAGCTTGATTTGATGACATA 360
Db 101 ArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 120
QY 361 CTTCAATGCCATGACATTGAGTTCGGTCTCTTGATCAGATTGTCAGTGGAACAATTCCT 420
Db 121 LeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePro 140
QY 421 GCTCTTCAGAAACTGAAACAGAGGGGAGACCCGGTTTCATTGGTATCACTGGTCTCCG 480
Db 141 AlaLeuGlnLysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGlyLeuPro 160
QY 481 TTAGATATTTTCATTATGTTCTGATCGAGTGCCTCCAGGACTGTCGATGTGATATG 540
Db 161 LeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeu 180
QY 541 TCATACTGTCATTACGGGTAAATCATTCGAGTTCGGTTCGGATTTACTACCTTACTTGAAG 600
Db 181 SerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLys 200
QY 601 AGCAAGGTGTGGGTGATAGTGTCTTCATAGCAATAGGCGCTCTTCACAGAACAA 660
Db 201 SerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGln 220
QY 661 GGTCTCTGTAATGGACCTGCTCCCTGAGCTCAAGTCTGCAAGCAAGCCGAGTT 720
Db 221 GlyProProGluTrpHisProAlaSerProGluLeuLysSerAlaSerLysAlaAlaVal 240
QY 721 GCTCACTGCAATCAAGGCAAGAGATCACAAAGTAGCTCTGCAATACAGATTTAGCA 780
Db 241 AlaHisCysLysSerLysGlyLysLysIleThrLysLeuAlaLeuGlnTyrSerLeuAla 260
QY 781 AACAAAGGAGATTTCGTGCGTGTGGTGGATGAGCTCTGCTCACAGTAGAAGAAAT 840
Db 261 AsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGluAsn 280
QY 841 GTTGACGAGTTACAGACTTGAAGTCTGGGAGTGGATGATCAAGAACTCTGCTGAGTT 900
Db 281 ValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluVal 300
QY 901 GAAGCTATTCTGAGCCTGTAAGAAATCTGACATGCCAAGTGGGAATCCATCAGAAC 957
Db 301 GluAlaIleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGlnAsn 319

RESULT 2

T28841
hypothetical protein F37C12.12 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T28841
R:Fulton, L.
A:Description: The sequence of C. elegans cosmid F37C12.
submitted to the EMBL Data Library, March 1994
A:Reference number: Z20530
A:Accession: T28841
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-439 <FULL>
A:Cross-references: UNIPROT:Q20127; EMBL:U00033; PIDN:AAC48300.1; GSPDB:GN00021; CESP:F37C12
A:Experimental source: strain Bristol N2; clone F37C12
C:Genetics:

A:Gene: CESP:F37C12.12
A:Map position: 3
A:Introns: 31/3; 63/3; 101/1; 127/1; 148/3; 204/3; 268/3; 312/3; 371/2
Alignment Scores:
Pred. No.: 7,41e-33 Length: 439
Score: 503.00 Matches: 120
Percent Similarity: 54.60% Conservative: 58
Best Local Similarity: 36.81% Mismatches: 106
Query Match: 30.43% Indels: 42
DB: 2 Gaps: 10

US-10-606-300-12 (1-960) x T28841 (1-439)

QY 7 AAAATAGAGCTTCGAGCTTTTGGGGAACACACAGGCTTAAGTTAGCGCGTTGGTTTGGT 66
Db 82 ArgMetAsnTyrArgGlnIleProGlyThrAspIleArgMetSerLysIleGlyPheGly 101
QY 67 GCCTCTCCGTCGGAAGTGTCTTCGTCACATGCCGAGATGATGCCGTCGCCACCGTG 126
Db 102 AlaAlaAlaIleGlyMetPheGlyAsnVal---GluAspSerIleIleLysIleVal 120
QY 127 CGCAGGCTTTCGCTCGGTATCAACTTCTTCGACACCTCCCGCTATTATGAGGAAACA 186
Db 121 GluThrAlaIleLysGlnGlyIleAsnTyrIleAspThrGlyTyrTyrSerGlnSer 140
QY 187 CTGTCTGAGAAATGCTTGGTAAGGACTAAAGGCTTTGCAAGTCCCTAGAGTGACTAC 246
Db 141 ArgSerGluSerIleLeuGlyLysAlaLeuSer-----LysIleProArgLysAlaTyr 158
QY 247 ATTGTGGCTACTAGTGTGGTAGA-----TATAAAGAAGGTTTGAATTTCACT 294
Db 159 TyrIleSerThrLysValGlyArgPheGluLeuAspTyrAlaArgThrPheAspPheArg 178
QY 295 GCTGAGAGAGTAAGAAAGAGATTGACGAGAGCTTGGAGAGCTTCACGCTTGATTTGTT 354
Db 179 AlaAspLysIleLeuGluSerLeuThrAsnSerLeuLysArgLeuGlnLeuThrTyrIle 198
QY 355 GACATA-----CTTCATTGCCATGACATTGAGTTCCGGTCTCTTTGATCAGATTGTG--- 405
Db 199 AspIleCysTyrValGlnIleHisAspAlaAspPheAlaProAsnGluSerIleValLeu 218
QY 406 AGTGAACAATCTGCTCTTCAGAACTGAAACAGAGGGGAGACCCGTTCACTTGGT 465
Db 219 TyrGluThrLeuGlnAlaLeuGluMetAlaLysSerSerGlyLysIleArgHisIleGly 238
QY 466 ATCACTGCTCTCCGTTAGAT-----ATTTTCACCTTATGTTCTTGATCGAGTG 513
Db 239 LeuThrGlyTyrProLeuGlyLysLeuValTyrSerPheArgTyrValTrpAsn----- 256
QY 514 CTTCCAGGACTGTCGATGTCATATTGTCATCTGTCATTACGGGCTTAATGATTCGAGC 573
Db 257 -----AlaProLysLeuTyrPheSerValAsnLeuSerIle 268
QY 574 TTGCTGGATTACTACCTTACTTGAAGAGCAAGGTGGGTGTGATGATGAGTCTTCTCCA 633
Db 269 -----ThrArgAsnIleAlaValIleAsnSerGlyAla 279
QY 634 TTGCAATGGGCTCTCTTACAGAACAAAGGCTCTCTGTAATGGCACCCCTGCTTCCCTGAG 693
Db 280 LeuCysTrpGlyLeuLeuThrGluLysGlyProProTyrHisProAlaSerAspGlu 299
QY 694 CTCAGTCTGCAAGCAAGCCGAGTTCCTCACTGCAAAATCAAGGGCAAGAGATCA 753
Db 300 IleLysGluAlaCysLeuAlaAlaThrThrTyrCysSerSerLysAsnIleSerIleSer 319
QY 754 AAGTTAGCTTGCATATACAGTTTAGCAACCAAGGAGATTTCGTCGGTGTGTTGGTGGGATG 813
Db 320 LysLeuAlaLeuAspTyrAlaLeuAsnPheProAsnValIleCysCysLeuValGlyMet 339
QY 814 AGCTCTGCTCACAGGTAGAGAAATGTT-----GCACAGATTACAGAG 858
Db 340 AspSerValGlnGlnValLeuAspAsnLeuGluLeuSerAsnPheSerArgIleThrAsp 359

QY	430	AAACTGAACACAGAGGGGAAGACCCGGTTCATTGGTATCTACATCGTCTTCGTTAGATATT	489
Db	165	LyseValArgAlaGluGlyValVallysaIaIleGlySerMetValSerAspAla	184
QY	490	TTCACTTATGTTCTTGATCGAGTGCCTCCAGGACCTGTCGATGTGATATTGTCATACGT	549
Db	185	LeuThrArgAla-----ValArgGluAlaAspLeuAspLeuIleMetValAlaGly	201
QY	550	CATTACGGCGTTAATGATTTCGACGTTGCTG-----GATTTCATTACCT	591
Db	202	ArgTyr-----ThrLeuLeuGluGlnProAlaAlaThrGluValLeuPro	216
QY	592	TACTTGAAGAGCAAAAGTGTGGGTGTGATAAGTCTTCATAGCAATGGCGCTCCTT	651
Db	217	AlaCysAlaGluAsnAlaThrGlyIleValAlaAlaSerValPheAsnSerGlyLeuLeu	236
QY	652	ACAGAAACAAGGTCCTCCT-----GAATGGCACCTGCTTCCCTCGTAGCTC	696
Db	237	AlaGlnSerGluProLysArgAspGlyArgTyrGluTyrGlyGlnLeuProAspGluLeu	256
QY	697	AGTCTGCAAGCAAAAGCCGAGTGTCTACTGCAATCAAAGGCGAAGAGATCACAAAG	756
Db	257	TrpAspArgLeuValArgIleAlaAlaIleCysArgAsnHisAspValProLeuProAla	276
QY	757	TTAGCTCTCGAATACAGTTTAGCAACACAGGAGATTTCGTCGCTGTTGGTGGGATGAGC	816
Db	277	AlaAlaIleGlnPheProLeuGlnSerAlaLeuValArgSerValValValGlyGlySer	296
QY	817	TCTGTCTCAGGTAGAGAAATGTTGCA-----GCAGTTACAGAGCTTGAAAGTCTGGG	873
Db	297	ArgProAlaGlnLeuThrGlnAsnAlaGluTyrAlaAlaLeuGluIleProAla--Gly	315
QY	874	ATGATCAGAAGAACTCTGCTGAGGTTGAAGCTATTCTCGAGCCT	918
Db	316	LeuTrpAlaGlu---LeuAlaGluAlaArgLeuIleProThrPro	329

RESULT 4

C83776

oxidoreductase BH1011 [imported] - Bacillus halodurans (strain C-125)

C/Species: Bacillus halodurans

C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004

C/Accession: NC3776

R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, Nucleic Acids Res. 28, 4317-4331, 2000

A/Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans

A/Reference number: AB3650; MUID:20512582; PMID:11058132

A/Status: preliminary

A/Accession: C83776

A/Molecule type: DNA

A/Residues: 1-297 <STO>

A/Cross-references: UNIPROT:Q9KE47; GB:AP001510; GB:BA000004; NID:g10173440; PI

A/Experimental source: strain C-125

C/Genetics:

A/Gene: BH1011

C/Superfamily: fission yeast pyridoxine 4-dehydrogenase

Alignment Scores:			
Pred. No.:	1,81e-21	Length:	297
Score:	362.50	Matches:	96
Percent Similarity:	52.43%	Conservative:	55
Best Local Similarity:	33.33%	Mismatches:	106
Query Match:	21.93%	Indels:	31
DB:	2	Gaps:	8

US-10-606-300-12 (1-960) x C83776 (1-297)

QY	19	CGAGCTTTGGGGAACACAGGCGTTAAGGTTAGCGCGTGTGGTTTGGTCTCTCCGCTC	78
Db	4	ArgGlnLeuGlyThrSerAspLeuHisValSerGluLeuGlyPheGlyCysMetSerLeu	23
QY	79	GGAAGTGTCTTCGGTCCAGTCGCGCGAAGATGATGCGCTCGCCACCGTGCAGGCGTTC	138

```
Db      24 GlyThr-----AspGluThrLysAlaArgArgIleMetAspGluValLeu 38
QY      139 CQTCTCGGTATCAACTTCTTCACACCTCCCGTATTATGGAGGAACACTGTCTCAGAAA 198
Db      39 GluLeuGlyIleAsnTyrLeuAspThrAlaAspLeuTyrAsnGlnGlyLeuAsnGluGln 58
QY      199 ATGCTTGTGAAGGACTAAGGCTTTGCAAGCTTCAGAGCTGCTAGAGTGAACATCTGTGCTACT 258
Db      59 PheValGlyLysAlaLeuLysGly-----ArgArgGlnAspIleIleLeuAlaThr 75
QY      259 AAGTCGTGT---AGATATAAAGAGT-----TTTGATTTCAGTCTGAG 300
Db      76 LysValGlyAsnArgPheGlnGlnGlyLysGluGlyTrpTrpAspProSerLysAla 95
QY      301 AGAGTAAGAAAGAGTATTCACAGAGCTTGGAGAGCTTCAGCTTGATTATGTTGACATA 360
Db      96 TyrIleLysGluAlaValLysAspSerLeuArgArgLeuGlnThrAspTyrIleAspLeu 115
QY      361 CTTCAATGCCATGACATTGAGTTCCGGTCTCTTGATCAGATGTGAGTGAACAATTCCT 420
Db      116 TyrGlnLeuHisGly-----GlyThrIleAspAspProIleAspGluThrIleGlu 132
QY      421 GCTCTTCAGAACTCAACAGAGGGAGACCCGGTTCATTGTGTATGCTGCTGCTCCG 480
Db      133 AlaPheGluGluLeuLysGlnGlyValIleArgTyrTyrGlyIleSerSerIleArg 152
QY      481 TTAGATATTTTTCACCTATGTTCTTCGATCGAGTGCCTCCAGGAGCTGTGCGATGTGATATTG 540
Db      153 ProAsnValIleLysGluTyrLeuLysArg-----SerAsnIleValSer 167
QY      541 TCATACTGTCAATACGGCGTTAATGATTCGACGTTGCTGGATTTACTACTTACTTTGAAG 600
Db      168 IleMetMetGlnTyrSerIleLeuAspArgArgProGluGluTrpPheProLeuIleGln 187
QY      601 AGCAAGGTGCGGTGATAGTCTCTCCATTAGCAATAGGAGGCGCTCTTACAGAACAA 660
Db      188 GluHisGlyValSerValValValArgGlyProValAlaArgGlyLeuLeuSerArgArg 207
QY      661 GGTCTCTCTGAA-----TGGCACCTCTGCTTCCCTCGAGCTCAAGTCTGCAAGC 708
Db      208 ProLeuProGluGlyGluGlyTyrLeuAsnTyrArgTyrAspGluLeuLysLeuArg 227
QY      709 AAAGCGCGAGTTGCTCACTGCGAAATCAAGGGCAAGAGATCACAAAGTTAGTCTGCAA 768
Db      228 GluSerLeu-----ProThrAspArgProLeuHisGluLeuAlaLeuGln 242
QY      769 TACAGTTTAGCAAAACAGAGAGATTTCGTGGTGTGGTTGGATGAGCTCTGCTCACAG 828
Db      243 TyrCysLeuAlaHisAspValValAlaThrValAlaAlaGlyAlaSerSerIleAspGln 262
QY      829 GTAGAAGAAAATGTTGCGAGAGTTT 852
Db      263 ValLysAlaAsnValGlnAlaVal 270

RESULT 5
S52890
hypothetical protein YMR041c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YMR041c
C:Species: Saccharomyces cerevisiae
C:Date: 08-May-1995 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C:Accession: S52890
R:Odell, C.; Bowman, S.
submitted to the EMBL Data Library, February 1995
A:Reference number: S52885
A:Accession: S52890
A:Molecule type: DNA
A:Residues: 1-335 <ODS>
A:Cross-references: UNIPROT:Q04212; EMBL:Z48502; NID:g695715; PIDN:CAA88407.1; PID:g6957
C:Genetics:
A:Gene: MIPS:YMR041c
A:Cross-references: SGD:S0004644
A:Map position: 13R
```

```
Alignment Scores:
Pred. No.: 4,3e-21 Length: 335
Score: 358.00 Matches: 97
Percent Similarity: 51.68% Conservative: 57
Best Local Similarity: 32.55% Mismatches: 112
Query Match: 21.66% Indels: 32
DB: 2 Gaps: 10

US-10-606-300-12 (1-960) x S52890 (1-335)

QY      124 GTGCGGAGGCTTTCGGTCTCGGTATCAACTTCTTCGACACCTCCCGTATTATGAGGA 183
Db      41 IleLysTyrAlaPheSerHisGlyLeuAsnAlaIleAspThrSerProTyrTyrGlyPro 60
QY      184 ACACCTGTCTGAGAAAATCGTTGTAAGGAGCTTAAAGGCTTTG-----CAAGTCCTCAGA 237
Db      61 -----SerGluValLeuTyrGlyArgAlaLeuSerAsnLeuAsnGluPheProArg 78
QY      238 AGTGACTACATTGGCTGCTACTAAGTGTGCTAGATAT---AAAGAAGGTTTTCATTCAGT 294
Db      79 AspThrTyrPheIleCysThrLysValGlyArgIleGlyAlaGluGluPheAsnTyrSer 98
QY      295 GCTGAGAGAGTAAGAAAGAGTATTGACGAGAGCTTGGAGAGCTTCAGCTTGATATGTT 354
Db      99 ArgAspPheValArgPheSerValHisArgSerCysGluArgLeuHisThrThrTyrLeu 118
QY      355 GACATACCTTCATTGCGCATGACATTGAGTTCCGGTCTCTTGATCAGATTTGTGAGTGAACA 414
Db      119 AspLeuValTyrLeuHisAspValGluPheValLysPheProAspIleLeu---GluAla 137
QY      415 ATTCCTGCTCTTCAGAACTGAAACAGAGGGGAGACCCGGTTCATTGCTGATCACTGGT 474
Db      138 LeuLysGluLeuArgThrLeuLysAsnLysGlyValIleLysAsnPheGlyIleSerGly 157
QY      475 CTTCCGTTAGATATTTTTCATTGCTTCTGATGAGTGCCTCA-----519
Db      158 TyrProIleAspPheIleThrTrpLeuAlaGluTyrCysSerThrGluGluSerAspIle 177
QY      520 GGGAGCTGCGATGATATTTGCTACTGCTTACCTACCGCGCTTAATGATTCGAGCTGCTG 579
Db      178 GlySerLeuAspAlaValLeuSerTyrCysAsnLeuAsnLeuGlnAsnAsnLysLeuLeu 197
QY      580 GATTACTACCTTACTTG-----AAGAGCAAGAGTGTGGGTGTGATGAAGTCTTCTCCA 633
Db      198 AsnPheArgGluArgLeuLeuArgAsnAlaLysLeuLysMetValCysAsnAlaSerIle 217
QY      634 TTAGCAATGGCGCTCTTACAGAACAGAGTCTCTCGAATGCGACCTGCTTCCCTGAG 693
Db      218 LeuSerMetSerLeuLeuArgSerGlnGluThrArgGlnPheHisProCysSerHisGlu 237
QY      694 CTCAGTCTGCAAGCAAGCCGCTGCTCTCACTGCAATCAAGGGCAAGAGATCACA 753
Db      238 LeuArgGluCysAlaSerGlnAlaLysTyrCysGlnGluGlnAsnValAspLeuAla 257
QY      754 AAGTAGCTCTGCAATACAGTTTACAGGTAGAGAAATGTTGACAGCAGTTCACAGAGCTT 810
Db      258 AspLeuAlaThrArgTyrAlaIle-----SerGluTrpValGlyLys 271
QY      811 -----ATGAGCTCTGCTCTCAGAGTAGAGAAATGTTGACAGCAGTTCACAGAGCTT 861
Db      272 GlyProValValLeuGlyValSerSerMetGluGluLeuLysLeuAlaLeuAspAsnTyr 291
QY      862 GAAAGTCTGGGGATGATCAAGAACTCTGCTGAGGTTGAAGCT---ATTCGAGCGCT 918
Db      292 GluIleValLysSerAsnGlyAsnArgLeuSerSerLysAspGlyGlnLeuValGluTyr 311
QY      919 GTAAAG-----RATCTGACATGCGCAAGTGGGATC 948
Db      312 IleGlnLysAsnIlePheLysGluHisPheAsnGluGluTrpSerSerGlyIle 329

RESULT 6
AF2990
oxidoreductase Atu3528 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
```

196 Glu1leAspCys1leLeuLeuAlaGlyAArgTyr-----ThrLeuLeuAsp 210
583 -----TTACTACCTTACTTTGAAGACCAAGGTGTGGGTGTGATAAGT 624
211 ArgSerAlaAlaAlaArgLeuLeuGlyAArgCysAlaGluThrGlyThrSerLeuValIle 230
625 GCTTCTCCATTAGCAATGGCCTCCCTTACAGAACACAGGTCCTCTCT-----GAA 672
231 GlyGlyValPheAsnSerGly1leLeuAlaThrGlyAlaLysProGlyAlaThrPheAsn 250
673 TGGCACCCCTGCTTCCCTCGAGCTCAAGTCTGCAAGCAAGCCGAGTGTCTACTGCAAA 732
251 TyrAsnGluAlaValProGluValMetGluArgValCysAlaMetGluAlaHisAlaAla 270
733 TCAAGGGCCGAAGATCATCAAAAGTTAGTCTTCCATATACATTTAGCAAAACAAGGAGATT 792
271 GlyHisGlyValAlaLeuAlaAlaAlaLeuHisPheProLeuGlnAsnThrAspVal 290
793 TCGTCGGTGTGTGGGATGAGCTCTGCTCTACAGGTAGAACAAAATGTTGCA----- 845
291 AlaSerValLeuIleGlyThrAlaLysProAspSerLeuArgAsnLeuSerIlePhe 310
847 -----GCAGTTACAGAGCTTGAAGTCTGGGGATGATCAAGAAACTCTGTCTGAG 897
311 GluThrAlaValProGlyAlaAlaTrpAlaGlyPheAspThrLeuAlaLeuGluAsp 329

RESULT 7
C98293
D-3-phospho-aldose 1-dehydrogenase (EC 1.1.1.122) [imported] - Agrobacterium tumefaciens
C/Species: Agrobacterium tumefaciens
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C/Accession: C98293
R:/Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Mar
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Mar
Science 294, 2323-2328, 2001
A./Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobact
A./Reference number: A97359; PMID:21608551; PMID:11743194
A./Status: preliminary
A./Accession: C98293
A./Molecule type: DNA
A./Residues: 1-329 <KUR>
A./Cross-references: UNIPROT:Q8U47; GB:AE007870; PIDN:AAK89869.1; PID:gl5159812
C/Genetics:
A./Gene: AGR L 2607
A./Map position: linear chromosome
C./Keywords: oxidoreductase

Alignment Scores:
Pred. No.: 3,69e-20 Length: 329
Score: 346.50 Matches: 102
Percent Similarity: 48.97% Conservative: 64
Best Local Similarity: 30.09% Mismatches: 120
Query Match: 20.96% Indels: 53
DB: 2 Gaps: 10

US-110-606-300-12 (1-960) x C98293 (1-329)

QY 10 ATAGAGCTTCGAGCTTTGGGGAACACAGGGCTTAAGGTTAGCGCGTGTGGTTGGTGCC 69
DB 1 MetClyThrArgLysLeuGlyAArgThrAlaLeuGluLeuSerGluLeuSerPheGlyAla 20
QY 70 TCTCCGCTCGGAAGTGTCTTCGCTCCAGTCCGCGAAGAGATGATCCGCTCGCCACCGTCGCG 129
DB 21 AlaGlyIleGlyAsnLeuTyrArgSerValSerArgGluAspAlaMetAlaThrLeuGln 40
QY 130 GAGGCTTTCGCTCGGTATCAACTTCTTCGACACCTCCCGTATATATGAGGAGAACACTG 189
DB 41 ThrAlaTrpAspAlaGlyIleAArgTyrPheAspThrAlaProTyrTyrGlyGlnGlyLeu 60
QY 190 TCTGAGAAATGCTTGGTAAGGACTAAAGGCTTTCGAGCTCCCTAGAGTGTACTACATT 249
DB 61 SerGluArgValGlyAspPheLeuGln-----GluLysProArgAspGluPheVal 78

Qy	250	GTGGCTACTAAGTGTGGTAGATATAAGAA	-----GATTTTCAGTCTCAGAGAGTAA	309
Db	79	LeuserThrLysValGlyArgIleLeuLysProAlaGluAlaGlyValThrProAspTyr	98	
Qy	280	GGTTT	-----GATTTTCAGTCTCAGAGAGTAA	309
Db	99	GlyPheValAspAlaLeuProPheIleValGluTyrAspTyrSerTyrAspGlyIleMet	118	
Qy	310	AAGAGTATTGACGAGAGCTTGGAGAGCGTTCACGTTGATTATGTACATATCTTCATTCG	369	
Db	119	ArgSerHisGluLeuSerLeuAlaArgLeuGlyLeuGlySerValAspIleLeuTyrVal	138	
Qy	370	CATGACATTGATGTCGGGTCTCTTCATCAG	-----ATTGTG	405
Db	139	HisAspLeuGluAlaThrThrLeuGlyGluGluAlaTyrArgHisHisPheGlyIlePhe	158	
Qy	406	AGTGAACA	---ATTCCCTGCTCTTCAGAACTCGAAACAGAGCGGGAAGACCCGGTTCATT	462
Db	159	ThrGluSerGlyIleGluAlaLeuHisGluLeuLysAlaLysGlyGlu	---Ile	175
Qy	463	GGTATCACTGGTCTTCCTCGGTAGATATTTTCACCTTATGTTCTTCGATCGAGTGTCCAGG	522	
Db	176	GlyAlaPheGlyLeuGlyValAsnGluValProAlaCysLeuAsnLeuMetGluIleAsp	195	
Qy	523	ACTGTCGATGATATTTGTCATCTGTCATTACGGCGTTAATGATTCGACGTTGCTGAT	582	
Db	196	GluIleAspCysIleLeuLeuAlaGlyArgTyr	-----ThrLeuLeuAsp	210
Qy	583	-----TTACTACCTTACTTGAAGCAAAAGTGTGGTGTGATAGT	624	
Db	211	ArgSerAlaAlaAlaArgLeuLeuGlyArgCysAlaGluThrGlyThrSerLeuValIle	230	
Qy	625	GCTTCTCCATTAGCAATGGCGCTCCTTCACAGAACAAAGTCTCCCT	-----GAA	672
Db	231	GlyGlyValPheAsnSerGlyIleLeuAlaThrGlyAlaLysProGlyAlaThrPheAsn	250	
Qy	673	TGGCACCTGCTTCCCTGAGCTCAAGTCTCGAAGCAAGCCGCGAGTGTCTCACTGCAAA	732	
Db	251	TyrAsnGluAlaValProGluValMetGluArgValCysAlaMetGluAlaHisAlaAla	270	
Qy	733	TCAAAGGCAAGACATCACAAGTTAGCTCTCGCAATACAGTTTACCAACAAGGATTT	792	
Db	271	GlyHisGlyValAlaLeuAlaAlaAlaLeuHisPheProLeuGlnAsnThrAspVal	290	
Qy	793	TCGTCCGTTGTGGTGGATGAGCTCTGCTCACAGGTAGAGAAATCTTCCA	-----	846
Db	291	AlaSerValLeuIleGlyThrAlaLysProAspSerLeuArgArgAsnLeuSerIlePhe	310	
Qy	847	-----GCAGTTACAGAGCTTGAAGAGTCTGGGATGGATCAAGAAACTCTGTCTGAG	897	
Db	311	GluThrAlaValProGlyValAlaAlaTropAlaGlyPheAspThrLeuAlaLeuGluAsp	329	

RESULT 8

H69966

conserved hypothetical protein yqkF - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004

C:Accession: H69966

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertaini, A.; Authors: Foulger, D.R.; Fritze, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallenbach, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.P.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Mauch, A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauel, V.; M. Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaei, V.; Sato, T.; Scanlon, A.; Authors: Schleif, S.; Schroeter, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Seron, aketuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

QY	10	ATAGAGCTTCGAGCTTTGGGGAACACAGCGCTTAAGGTTAGCGCGCGTGTGTTTGGTGCC	99
DB	2	ValAspTyrArgTyrLeuGlyArgSerAlaLeuLysValSerProLeuSerLeuGlyThr	21
QY	70	TCTCGCTCGGAAGTGCTCTTCGGTCCAGTCGCCGAAGATGATGCGTCGCCACCGTGGC	129
DB	22	MetMetPheGly-----GlyProThrProAspValAlaTyrArgIleIleAsp	38
QY	130	GAGGCTTTCGGTCTCGGTAACTTCCTTCGACACCTCCCGCTATTATTATGGAGGAACACTG	189
DB	39	LysAlaArgGluGlnGlyIleAsnPheIleAspThrAlaAspValTyrHisAspGlyLys	58
QY	190	TCTGAGAAAATCGTTGGTAAGGACTAAAGCTTTGCAAGTCCCTACAGTGACTACTCATT	249
DB	59	SerGluGluValValGlyArgGlyIleLysAla-----SerArgAspHisTrpVal	75
QY	250	GTGGCTACTAAG---TGTTGTAGATATAAGAGAAGCTTTGTATTTC-----AGTGCT	297
DB	76	LeuAlaThrLysPheValAsnSerHisThrLysGlyProAsnLeuGlyGlyHisSerArg	95
QY	298	GAGAGATTAAGAAAGAGATTTCAGCAGAGCTTGAGAGGCTTCAGCTTGATTATGTTGAC	357
DB	96	LysTrpValIleGluThrValGluAsnSerLeuArgArgLeuAsnThrAspTyrIleAsp	115
QY	358	ATACTTCATTGTCATCACATTGAGTTTCGGTCTCTTGTATCAGATTGTGAGTGAACAATT	417
DB	116	IleLeuTyrPheHisArgAlaValPhe-----AspAlaProLeuGluGluProVal	132
QY	418	CTGCTCTTTCAGAAACTGAGAAACAAGAGGGAGACCCGGTTTCATTGCTATCACT---	471
DB	133	ArgAlaIleAlaAspLeuIleArgAlaGlyLysLeuArgTyrPheGlyValSerAsnPhe	152
QY	472	---GGTCTTCGTAGATATTTCCTACTTATGTTCTTGATGAGTGCCTCCAGGAGCTGC	528
DB	153	ArgGlyTrpArgIleAlaGluIleSerHisLeuAlaAspGlnLeu-----GlyIleAsp	170
QY	529	GATGTGATATTGTCATCACTGTCATTACGCG---GTTAATGATTCCGAGCTTGCTGGATT	585
DB	171	ArgProValAlaSerGlnProLeuTyrAsnIleValAsnArgThrAlaGluAlaGluGln	190
QY	586	CTACCTTACTTGAAGAGCAAGAGTGTGGTGTGATAAGTGTCTTCCATTAGCAATGGGC	645
DB	191	LeuProAlaAlaAsnHisTyrGlyLeuGlyValValSerTyrSerProLeuAlaArgGly	210
QY	646	CTCCTTACAGAACAGGTCTCTCT-----	669
DB	211	ValLeuThrGlyLysTyrGlnProGlyGluGlnProGlyAlaAspThrArgValGlyArg	230
QY	670	-----CAATGGCACCTGCTTCCCTCGAGCTCAAGTCTGCA	705
DB	231	GlyAspLysArgValLeuGluThrGluTrpArgProGluSerValGluIle-----Ala	248
QY	706	AGCAAGCCGCGAGTTGCTCACTGCAAAATCAAGAGGCAAGAGATCACAAAATTAGTCTGT	765
DB	249	GlnLysValAla---AlaHisAlaAlaSerLysGlyValSerAlaAlaAspPheAlaLeu	267
QY	766	CAATCAGTTTAGCAAAACAGAGAGATTTCGTGGTGTGTTGGATGGAGCTCTGCTCA	825
DB	268	AlaTrpValLeuAsnAsnLysPheValThrAlaAlaIleThrGlyProArgThrGluGlu	287
QY	826	CAGGTAGAAGAAAATGTTTCGACAGCTTACAGAGCTTGAAAGCTCTGGGGAGTGATCAAGA	885
DB	288	HisTrpAsnSerTyrValArgAla-----LeuAspValLys	299
QY	886	ACTCTCTGTGAGTTGAAGCTATTCTCGAGCCTGTAAAGAAATCGACATGCCCAAGT	942
DB	300	IleAspAlaGluAspGluAlaLeuValAspSerLeuValThrThrGlyHisProSer	318

RESULT 11
H90938
hypothetical protein ECs2480 [imported] - Escherichia coli (strain O157:H7, substrain RM
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence revision 18-Jul-2001 #text change 09-Jul-2004

C;Accession: H90938
F;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
G;Sawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
H;DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gen
A;Reference number: A9629; MUID:21156231; PMID:11258796
A;Accession: H90938
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-326 <HAY>
A;Cross-references: UNIPROT:Q8XDV3; GB:BA000007; PIDN:BAB35903.1; PID:gl3361947; GSPDB:C
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
C;Gene: ECs2480
C;Superfamily: fission yeast pyridoxine 4-dehydrogenase

Alignment Scores:
Pred. No.: 1.61e-17 Length: 326
Score: 314.00 Matches: 103
Percent Similarity: 43.44% Conservative: 46
Best Local Similarity: 20.03% Mismatches: 128
Query Match: 19.00% Indels: 66
DB: 2 Gaps: 12

US-10-606-300-12 (1-960) x H90938 (1-326)

Qy 1 ATGACGAATAAGACTTCGACCTTTGGGGAACACAGGCTTAAGGTTAGCGCGTTGGT 60
Db 1 MetLysLysIle-----ProLeuGlyThrAspIleThrLeuSerArgMetGly 17
Qy 61 TTTGTGCGCTCCCGCTCGGAAGTGTCTCGGTCCAGTCCGCCGAAGATGAT----- 111
Db 18 LeuGlyThrTrpAlaIleGlyGly-----GlyProAlaTrpAsnGlyAspLeuAspArg 35
Qy 112 -----GCCGTGCCACCGTCGCGGAGGCTTTCCGTCTCGGTATCACTTCTTCGACACC 165
Db 36 GlnIleCysIleAspThrIleLeuGluAlaHisArgCysGlyIleAsnLeuIleAspThr 55
Qy 166 TCCCGGTATTATGGAGGAACACTGTCGTGAGAAAACTCTGTGTAAGGACTAAAGCGCTTTG 225
Db 56 AlaProGlyTyrAsnPheGlyAsnSerGluValIleValGlyGlnAlaLeuLysLysLeu 75
Qy 226 CAAGTCCTTAGAAGTGACTACATTGTGGCTACTAAGTGTGGT----- 267
Db 76 -----ProArgGluGlnValValGluValGluThrLysCysGlyIleValTrpGluArgLys 93
Qy 268 -----AGATATAAAGAAAGGTTTTTATTGTTCCAGT 294
Db 94 GlySerLeuPheAsnLysValGlyAspArgGlnLeuTyrLys-----AsnLeuSer 110
Qy 295 GCTGAGAGAGTAAGAAAGATATTGACGAGAGCTTCGAGAGGCTTCAGCTTGATTATGTT 354
Db 111 ProGluSerIleArgGluGluValGluAlaSerLeuGlnArgLeuGlyPheAspTyrIle 130
Qy 355 GACATACCTTCATTCGATGACATTAAGTTCGGGTCTCTTCATCAGATTGTCAGTGAACA 414
Db 131 AspIleTyrMetThrHisTrpGlnSerValProPheTyrThrProIleAlaGluThr 150
Qy 415 ATTCCTGCTCTTCAGAAACTGAAACAAAGAGGGGAAGACCCGGTTTCATTGGTATCACTGGT 474
Db 151 ValAlaValLeuAsnGluLeuLysAlaGluGlyLysIleArgAlaIleGlyAlaAsn 170
Qy 475 CTTCCGTTAGATATTTTCACTTATTGTCGAGTCGCTCCAGGACTGTCGATGTG 534
Db 171 ValAspAlaAsp-----HisIleArgLysLeuGlnTyrGlyGluLeuAspIle 187
Qy 535 ATATTGTCTACTGTCTCATTCACGCGCTTAATGATTCGACGTTG---CTGGATTTACTACT 591
Db 188 Ile-----GlnAlaLysTyrSerIleLeuAspArgAlaMetGluAsnGluLeuPro 205
Qy 592 TACTTTGAAGCAAGGTGTGGTGTGATAAGTGCTTCTCCATTAGCAATGGGCTCTT 651
Db 206 LeuCysArgAspAsnGlyIleValGlnValTyrSerProLeuGluGlnGlyLeuLeu 225

```
QY 652 ACAGAACAAAGTCTCTCTGAATGAGCACCCTGCTTCCCTGAGCTCAAGTCTGCAAGCAA 711
|||
Db 256 ThrGlyThrIleThrArgAspTyrValProGlyGlyAlaArg-----AlaAsnLys 242
|||
QY 712 GCCGAGTTGCTCAC----- 726
|||
Db 243 ValTrpPheGlnArgGluAsnMetLeuLysValIleAspMetLeuGluGlnTrpGlnPro 262
QY 727 ---TGCAAAATCAAAGGCGAAGAGATCAAAAGTTAGCTCTGCAATACAGATTTAGCAAAAC 783
|||
Db 263 LeuCysAlaArgTyrGlnCysThrIleProThrLeuAlaLeuAlaTrpIleLeuLysGln 282
QY 784 AAGGAGATTTTCGCTGCTGTTGGTGGAGAGCTCTGCTCAAGTGAAGAAATGTT 843
|||
Db 283 SerAspLeuIleSerIleLeuSerGlyAlaThrAlaProGluGlnValArgGluAsnVal 302
QY 844 GCAGCAGTTACAGAGCTTGAAGTCTGGGATCGGATCAAGAACTCTCTCTGAGGTTGAA 903
|||
Db 303 AlaAlaLeu-----AsnIleAsnLeuSerAspAlaAsp 313
QY 904 GCTATTCTC 912
|||
Db 314 AlaThrLeu 316

RESULT 12
D85787
hypothetical protein 22809 [imported] - Escherichia coli (strain O157:H7, substrain EDL9
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: D85787
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: D85787
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-326 <STO>
A:Cross-references: UNIPROT:O8XDV3; GB:AE005174; NID:G12515802; PIDN:AAG56760.1; GSPDB:C
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
C:Gene: 22809
C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Alignment Scores:
Pred. No.: 1.61e-17 Length: 326
Score: 314.00 Matches: 103
Percent Similarity: 43.44% Conservative: 46
Best Local Similarity: 30.03% Mismatches: 128
Query Match: 19.00% Indels: 66
DB: 2 Gaps: 12

US-10-606-300-12 (1-960) x D85787 (1-326)
QY 1 ATGCGAAATAGAGCTTCGAGCTTTGGGACACAGGGCTTAAGGTTAGCGCCGTTGTT 60
|||
Db 1 MetLysLysIle-----ProLeuGlyThrThrAspIleThrLeuSerArgMetGly 17
QY 61 TTTGGTGCCTCTCCGCTCGGAAGTCTCTTCGGTCCAGTCGCGGGAAGATGAT 111
|||
Db 18 LeuGlyThrTrpAlaIleGly-----GlyProAlaTrpAsnGlyAspLeuAspArg 35
QY 112 -----GCCGTGCCAGCGTGGCGAGGCTTCGCTCTCGGTATCACTTCTTCACACC 165
|||
Db 36 GlnIleCysIleAspThrIleLeuGluAlaHisArgCysGlyIleAsnLeuIleAspThr 55
QY 166 TCCCGGTATTAGGAGGAACACTGTCTGAGAAATGCTTGGTAAGGGACTAAAGGCTTTG 225
|||
Db 56 AlaProGlyTyrAsnPheGlyAsnSerGluValIleValGlyGlnAlaLeuLysLysLeu 75
QY 226 CAAAGTCCCTAGAGTGAAGTACTACATTGCTGGCTACTAAGTGTGGT----- 267
```

```
Db 76 -----ProArgGluGlnValValGluThrLysCysGlyIleValTrpGluArgLys 93
|||
QY 268 -----AGATATAAAGAAAGTTTTCATTGATTCAGT 294
|||
Db 94 GlySerLeuPheAsnLysValGlyAspArgGlnLeuTyrLys-----AsnLeuSer 110
QY 295 GCTGAGAGATTAAGAAAGATTTGACGAGAGCTTTGGAGAGCTTCAGCTTGATTTGTT 354
|||
Db 111 ProGluSerIleArgGluGluValGluAlaSerLeuGlnArgLeuGlyPheAspTyrIle 130
QY 355 GACATACCTTCATGCCATGACATTCAGTTCGGGTCTCTTGATCAGATTCGTGATGAACA 414
|||
Db 131 AspIleTyrMetThrHisTrpGlnSerValProProPheTyrThrProIleAlaGluThr 150
QY 415 ATTCTGCTCTTCAGAAACTGAACACAGAGGGAGACCCGGTTCATTGGTTCATCTGTT 474
|||
Db 151 ValAlaValLeuAsnGluLeuLysAlaGluGlyIleArgAlaIleGlyAlaAlaAsn 170
QY 475 CTTCCGTTAGATATTTTCATCTTATGTTGATCGAGTGCCTCCAGGAGCTGTCGATG 534
|||
Db 171 ValAspAlaAsp-----HisIleArgGluTyrLeuGlnTyrGlyGluLeuAspIle 187
QY 535 ATATTGTCATCTGTCATTACGGCGTTAATGATTCGACGTTG---CTGGATTTTACTACT 591
|||
Db 188 Ile-----GlnAlaLysTyrSerIleLeuAspArgAlaMetGluAsnGluLeuPro 205
QY 592 TACTTGAAGACAAAGGTGCGGTGTGATAAGTGTCTTCCATTAGCAATGGCCCTCTT 651
|||
Db 206 LeuCysArgAspAsnGlyIleValGlnValTyrSerProLeuGluGlnGlyLeuLeu 225
QY 652 ACAGAACAAAGTCTCTCTGAATGAGCACCCTGCTTCCCTGAGCTCAAGTCTGCAAGCAA 711
|||
Db 226 ThrGlyThrIleThrArgAspTyrValProGlyGlyAlaArg-----AlaAsnLys 242
QY 712 GCCGAGTTGCTCAC----- 726
|||
Db 243 ValTrpPheGlnArgGluAsnMetLeuLysValIleAspMetLeuGluGlnTrpGlnPro 262
QY 727 ---TGCAAAATCAAAGGCGAAGAGATCAAAAGTTAGCTCTGCAATACAGATTTAGCAAAC 783
|||
Db 263 LeuCysAlaArgTyrGlnCysThrIleProThrLeuAlaLeuAlaTrpIleLeuLysGln 282
QY 784 AAGGAGATTTTCGCTGCTGTTGGGATGAGCTCTCTCTCAGAGTGAAGAAATGTT 843
|||
Db 283 SerAspLeuIleSerIleLeuSerGlyAlaThrAlaProGluGlnValArgGluAsnVal 302
QY 844 GCAGCAGTTACAGAGCTTGAAGTCTGGGATCGGATCAAGAACTCTCTCTGAGGTTGAA 903
|||
Db 303 AlaAlaLeu-----AsnIleAsnLeuSerAspAlaAsp 313
QY 904 GCTATTCTC 912
|||
Db 314 AlaThrLeu 316
```

```
RESULT 13
C64937
hypothetical protein b1771 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
C:Accession: C64937
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: C64937
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-326 <BLAT>
A:Cross-references: UNIPROT:P77256; GB:AE000272; GB:U00096; NID:G1788067; PIDN:AAC74841
A:Experimental source: strain K-12, substrain MGL655
C:Superfamily: fission yeast pyridoxine 4-dehydrogenase
```

Alignment Scores:

Pred. No.: 1,94e-17 Length: 326
Score: 313.00 Matches: 103
Percent Similarity: 43.44% Conservative: 46
Best Local Similarity: 30.03% Mismatches: 128
Query Match: 18.94% Indels: 66
DB: 2 Gaps: 12

US-10-606-300-12 (1-960) x C64937 (1-326)

```
QY 1 ATGACGAAATAGAGCTTCGACCTTTGGGACACACAGGCTTAAGTTAGCGCCCTGGT 60
DB 1 MetLysLysIle-----ProLeuGlyThrAspIleThrLeuSerArgMetGly 17
QY 61 TTTGGTGCCTCCGCTCGAAGTGTCTTCGGTCCAGTCGCCGAAGATCAT----- 111
DB 18 LeuGlyThrTrpAlaIleGlyGly-----GlyProAlaTrpAsnGlyAspLeuAspArg 35
QY 112 -----GCCGTGCCACCGTCGCGAGGCTTTCCGTCTCGGTATCAACTTCTTCACACC 165
DB 36 GlnIleCysIleAspThrIleLeuGluAlaHisArgCysGlyIleAsnLeuIleAspThr 55
QY 166 TCCCGTATTATGGAGGACACTGCTCGAATAATGCTTGGTAAGGACTAAGGCTTGT 225
DB 56 AlaProGlyTyAsnPheGlyAsnSerGluValIleValGlyGlnAlaLeuLysLeu 75
QY 226 CAAGTCCCTAGAGTACACTACATTGTGGCTACTAAGTGTGGT----- 267
DB 76 -----ProArgGluGlnValValGluThrLysCysGlyIleValTrpGluArgLys 93
QY 268 -----AGATATAAAGAAGGTTTTCATTTCAAGT 294
DB 94 GlySerLeuPheAsnLysValGlyAspArgGlnLeuTyLys-----AsnLeuSer 110
QY 295 GCTGAGAGAGTAAGAAGAGTATTACGAGAGCTTGGAGAGGCTTCAGCTTGATTATGT 354
DB 111 ProGluSerIleArgGluGluValAlaAlaSerLeuGlnArgLeuGlyIleAspThrIle 130
QY 355 GACATACITTCATTCGCATGACATTCAGTTTCGGGTCTCTTTGATCAGATTGTGAGTAAACA 414
DB 131 AspIleTyMetThrHisTrpGlnSerValProProPheThrProIleAlaGluThr 150
QY 415 ATTCCTGCTCTCAGAACTGAAACAGAGGGGAGAGCCGGTTTCATTTGATATCATGGT 474
DB 151 ValAlaValLeuAsnGluLysSerGluGlyLysIleArgAlaIleGlyAlaAlaAsn 170
QY 475 CTTCGGTTAGATATTTCATCTTATGTTCTTGATCGAGTCCCTCCAGGAGCTCGCATNG 534
DB 171 ValAspAlaAsp-----HisIleArgGluTyLeuGlnTyGlyGluLeuAspIle 187
QY 535 ATATTGTATCATCTGTATTCACCGCTTAATGATTCGACGTTG---CTGGATTACTACCT 591
DB 188 Ile-----GlnAlaLysTySerIleLeuAspArgAlaMetGluAsnGluLeuLeuPro 205
QY 592 TACTTGAAGACGAAGGTGGGTGTGATAAGTCTTCATTTAGCAATGGGCTCCTTT 651
DB 206 LeuCysArgAspAsnGlyIleValValGlnValTySerProLeuGluGlnGlyLeuLeu 225
QY 652 ACAGAACAGAGTCCCTCCGAATGGCACCTGCTTCCCTGAGCTCAAGTCTCAAGCAAA 711
DB 226 ThrGlyThrIleThrArgAspTyValProGlyGlyAlaArg-----AlaAsnLys 242
QY 712 GCCGCGATTGCTCAC----- 726
DB 243 ValTrpPheGlnArgGluAsnMetLeuLysValIleAspMetLeuGluGlnTrpGlnPro 262
QY 727 -----TGCMAATCAAGGGCAAGAGATCACAAAGTTAGCTCTCGAATACAGTTTACGAAC 783
DB 263 LeuCysAlaArgTyGlnCysThrIleProThrLeuAlaLeuAlaTrpIleLeuLysGln 282
QY 784 AAGGAGATTTCGTGCGTGTGGTGGATGAGCTCTGTCTCACAGTAGAAGAAATGTT 843
```

```
DB 283 SerAspLeuIleSerIleLeuSerGlyAlaThrAlaProGluGlnValArgGluAsnVal 302
QY 844 GCAGCAGTTACAGAGCTTTGAAAGTCTGGGATCGGATCAAGAAACTCTCTGTGAGGTGAA 903
DB 303 AlaAlaLeu-----AsnIleAsnLeuSerAspAlaAsp 313
QY 904 GCTATTCTC 912
DB 314 AlaThrLeu 316
RESULT 14
H83427
probable oxidoreductase PA1739 [imported] - Pseudomonas aeruginosa (strain PAO1)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C:Accession: H83427
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic path
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: H83427
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-323 <STO>
A:Cross-references: UNIPROT:Q91228; GB:AE004600; GB:AE004091; NID:g9947712; PIDN:AAG051;
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: PA1739
C:Superfamily: fission yeast pyridoxine 4-dehydrogenase
Alignment Scores:
Pred. No.: 7.9e-17 Length: 323
Score: 305.50 Matches: 95
Percent Similarity: 46.63% Conservative: 57
Best Local Similarity: 29.14% Mismatches: 127
Query Match: 18.48% Indels: 47
DB: 2 Gaps: 11
US-10-606-300-12 (1-960) x H83427 (1-323)
```

```
QY 25 TTGGGACACACAGGCTTAAGTTAGCGCGCTTGGTTGGTCTCCGCTCGGAAGT 84
DB 6 LeuGlyAsnSerGlyLeuLysValSerArgLeuCysLeuGlyCysMetThrTyArgLeu 25
QY 85 ---GTCTTCGGTCCA-----GTCCCGAAGATGATGCGTCGCCACCGTCGCCGAGGCT 135
DB 26 ProAlaTrpArgProTrpValLeuAspGluGluArgAlaArgProPheIleArgGluAla 45
QY 136 TTCGCTCGGTATCAACTTCTTCACACCTCCCGTATTTATGGAGGAACACTGTCTGAG 195
DB 46 LeuGluAlaGlyIleAspPheAspSerAlaAspIleTySerThrGlyGluSerGlu 65
QY 196 AAAATGCTTGTAAAGGACTAAAGCTTTGCAAGTCCCTAGAGTGACTACATTGTGGCT 255
DB 66 ArgIleLeuGlyArgAlaLeuArgAspPhe---AlaGlnArgGluAspValIleAla 84
QY 256 ACTAAG-----TGTGGTAGATATAAAGAAGT 282
DB 85 ThrLysAlaPhePheProMetSerAspArgProAsnAlaCysGly----- 99
QY 283 TTTGATTTTCAGTGTGAGAGATGAAGAGTATTATGACGAGCTTGGAGAGGCTTCAG 342
DB 100 -----LeuSerArgLysHisLeuLeuAlaSerValAspAlaSerLeuArgArgLeuGly 117
QY 343 CTTGATTATGTTGACATCTTCATTGCGCATGACATTGAGTTTCGGGTCTCTTCATCAGATT 402
DB 118 ThrAspTyLeuAspPheValIleHisArgPheAsp-----ProAspThrPro 134
QY 403 GTGAGTGAACAATTCCTGCTCTTCAGAACTGAACAAGGGGGAAGCCGGTTCATT 462
DB 135 IleGluGluThrCysGluThrLeuAspSerLeuValArgValGlyLysValArgTyLeu 154
```

```
Oy 463 GGTATCACTGCTGCTTCGTTAGATATTTCTACTTATGTTCTTGAT---CGAGTGCCTCCA 519
Db 155 GlyAlaSerSerMetProAlaTrpArgPheMetLysMetLeuAlaPheGlnArgHis 174
Oy 520 GGGACTGCGATGATATTCATCATCTGCTCATTTACGGCGTT-----AATGATTCG 570
Db 175 GlyLeuAlaGlnPheIleSerMetGlnSerGlnInTrpAsnLeuIleValArgGluAspGlu 194
Oy 571 ACCTTGCTGGATTACTTACTTACTTGAAGAGCAAGGTGGGTGCTGATAGTCTTCT 630
Db 195 Glu-----AspLeuValProLeuCysArgGluGluGlyIleAlaLeuThrProTrpSer 212
Oy 631 CCATTAGCAATGGGCTCCTT-----ACA 654
Db 213 ProLeuAlaArgGlyLeuLeuAlaGlyAlaArgSerAlaGlyThrLeuArgThrArgThr 232
Oy 655 GAACAAGGTCTCTGTAATGACCCCTCTTCCCTGAGCTCAAGTCTGCAAGCAAGCC 714
Db 233 AspGluGlnAlaProArgTrpTrpGlyArgGluGluValGluSerThrLeuGlyAla 252
Oy 715 GCAGTTGCTCACTGCAAAATCAAAAGGGCAAGAGATCACAAAGTTAGCTCTGCAATACAGT 774
Db 253 LeuGluLysLeuAlaAlaArgGlyLeuProAlaGlnLeuAlaLeuAlaTrpLeu 272
Oy 775 TTAGCAACAAGGAGATTTCGCTGGTGTGGTGGATGAGCTCTGCTCACAGGTAGAA 834
Db 273 LeuGlyArgAsnGlyValAlaAlaProIleValGlyLeuSerArgProHisLeuGlu 292
Oy 835 GAAATGTTGACGAGTTACAGAGTCTGAAAGTCTGGGGATGGATCAAGAACTCTGTCT 894
Db 293 AspAlaLeuAlaLeuThr-----LeuAspLeuAlaGluGluCysAla 308
Oy 895 GAGGTTGAAGCTATTCTC 912
Db 309 ThrLeuGluAlaProLeu 314
```

RESULT 15
AB0808
probable ion-channel protein STY2647 [imported] - Salmonella enterica subsp. enterica serovar Typhi
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: This species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AB0808
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-332 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD07644.1; PID:G16503631; GSPDB:GN00176
C:Genetics:
A:Gene: STY2647
C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Alignment Scores:
Pred. No.: 7.94e-17 Length: 332
Score: 305.50 Matches: 101
Percent Similarity: 50.76% Conservative: 65
Best Local Similarity: 20.89% Mismatches: 126
Query Match: 18.48% Indels: 35
DB: 2 Gaps: 12

US-10-606-300-12 (1-960) x AB0808 (1-332)

```
Oy 10 ATAGAGCTTCGAGCTTTGGGCAACACAGGCTTAAGGTTAGCGCGTGGTTGGTGCC 69
Db 13 MetGluTrpArgCysGlyArgSerGlyValLysLeuProAlaIleSerLeuGly--- 31
```

Search completed: November 13, 2005, 08:31:13
Job time : 34.5 secs

```
Oy 70 TCTCCGCTCGAAGTGTCTTGGTCCAGTCCCGAA---GATGATGCGTCCGCCACCGTG 126
Db 32 -----LeuTrpHisAsnPheGlyAspThrArgValGluAsnSerArgAlaLeuLeu 49
Oy 127 CGCAGGCTTCCGCTCGGTATCACTTCTTCACACCTCCCGCTATATATGA----- 180
Db 50 GlnArgAlaPheAspLeuGlyIleThrHisPheAspLeuAlaAsnAsnTrpGlyProPro 69
Oy 181 ---GGAACACACTGTCTGAGAAATCTCTGTAAGGACTTAAAGGCTTTTCAAGTCCCT--- 234
Db 70 ProGlySer---AlaGluCysAsnPheGlyArgIleLeuGlnGluAspPheLeuProTrp 88
Oy 235 AGAAGTCACTACATTTGGTGGTCTAAGTGTGTGTAGATATAAAGAGGTTTGTATTCAGT 294
Db 89 ArgAspGluLeuValIleSerThrLysAlaGlyTrpMetTrpAspGlyProTrpGly 108
Oy 295 GCTGAGAGAGTAAGAAAG-----AGTATTGACGAGAGCTTGAGAGGCTTCAG 342
Db 109 AspTrpGlySerArgLysTrpLeuIleAlaSerLeuAspGlnSerLeuLysArgMetGly 128
Oy 343 CTTGATTATGTGACATCTTCACTGTCATGACATTCGAGTTCGGGTCTCTTGTGATCAGATT 402
Db 129 LeuGluTrpValAspIlePheTrpHisArgProAsp-----ProGluThrPro 145
Oy 403 GTGAGTGAACAATTCCTGCTCTTTCAGAAAACAGAGGGAAGACCCGTTTCATT 462
Db 146 LeuLysGluThrMetLysAlaLeuAspHisLeuValArgHisGlyLysAlaLeuTrpVal 165
Oy 463 GGTATCACTGCTTCGTTAGATATTTCTTCTTCTTCTGATCGAGTCCCTCCAGG 522
Db 166 GlyIleSerAsnTrpProAlaAspLeuAlaArgGlnAlaIleAspIleLeuGluAspLeu 185
Oy 523 ACTGTCGATGATATTTGTCATCTGTCATTCGCGGTTAATGATTCGACCTGCTGGAT 582
Db 186 GlyThrProCysLeuIleHisGlnProLysTrpSerLeuPheGluArgTrpValGluAsp 205
Oy 583 ---TTACTACTTACTTTGAAGAGCAAGGTGGTGGTGTGATGATGCTTCTTCCATTAGCA 639
Db 206 GlyLeuLeuThrLeuLeuGlnGluLysGlyValGlySerIleAlaPheSerProLeuAla 225
Oy 640 ATGGGCTCTCTTACAGAACAA-----GGTCTCTGTAATGGCACCCTGCTTCC--- 687
Db 226 GlyGlyGlnLeuThrAspArgTrpLeuAsnGlyIleProGluAspSerArgAlaAlaSer 245
Oy 688 -----CCTGAGCTCAAGTCTCAAGCAAGCGCGAGTTGCTCAC--- 726
Db 246 GlySerArgPheLeuLysProGluGlnIleThrAlaAspLysLeuGluLysValArgArg 265
Oy 727 -----TGCAAAATCAAGGCAAGAGATCAAAAGTTAGTCTTCTGCAATACAGT 774
Db 266 LeuAsnGluLeuAlaAlaArgGlyGlnLysLeuSerGlnMetAlaLeuAlaTrpVal 285
Oy 775 TTAGCAACAAGGAGATTTCGCTGGTGTGGTGGATGAGCTCTGCTCACAGGTAGAA 834
Db 286 LeuArgAsnAspAsnValThrSerValIleGlyAlaSerLysProSerGlnIleGlu 305
Oy 835 GAAATGTTGACGAGCTTACAGAGCTTCAAGTCTGGGGATGGATCAAGAAACTCTGTCT 894
Db 306 Asp-----AlaValGlyMetLeuAlaAsnArgArgPheSerAlaAlaGluCysAla 322
Oy 895 GAGTTGAAGCTATTCTCGAG 915
Db 323 GluIleAspAlaIleLeuGlu 329
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 13, 2005, 05:01:53 ; Search time 4375 Seconds
(without alignments)
10632.459 Million cell updates/sec

Title: US-10-606-300-12
Perfect score: 960
Sequence: 1 atgcgaaatagagcttcg.....gtggaatccatcagaactaa 960

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_to.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	960	100.0	960	6	AR405970
2	960	100.0	960	6	AX259907 Sequence
3	960	100.0	960	6	AX374959 Sequence
4	960	100.0	960	8	AJ417563 Arabidops
5	960	100.0	960	8	AY090337 Arabidops
6	960	100.0	1092	8	AY050377 Arabidops
7	582.4	60.7	1215	8	AY176585 Actinidia
8	570.2	59.4	1221	8	AY264803 Malus x d
9	518.8	54.0	1239	8	AB160990 Spinacia
10	454	47.3	1234	8	AK060431 Oryza sat
11	452.4	47.1	1305	8	AK067039 Oryza sat
12	450.8	47.0	1275	8	AK102223 Oryza sat
13	435	45.3	98124	8	ATT1611 Arabidops
14	435	45.3	192861	8	AL161584 Arabidops
15	435	45.3	199536	8	ATCHRIV79 Arabidops
16	217	22.6	360	11	CF384305 Arabidops
17	190.2	19.8	118803	8	AC124214 Medicago
18	173	18.0	142246	8	CNS08C8B Oryza sat
19	151.2	15.8	303446	1	AE016941 Bacteroid

20	134.4	14.0	2105	3	AK112151	AK112151 Clona int
c 21	128.6	13.4	311050	1	BX294133	BX294133 Pirellula
22	110	11.5	1038	6	CQ600264	CQ600264 Sequence
23	104.8	10.9	110000	2	AP006493_5	Continuation (6 of
24	98.2	10.2	1051	6	CQ612720	CQ612720 Sequence
25	98.2	10.2	1191	3	AY118797	AY118797 Drosophil
26	80.2	8.4	885	6	CQ579387	CQ579387 Sequence
c 27	71.2	7.4	3724	6	CQ600263	CQ600263 Sequence
c 28	71.2	7.4	20956	2	AC018265	AC018265 Drosophil
c 29	71.2	7.4	182182	3	AC007889	AC007889 Drosophil
c 30	71.2	7.4	226076	3	AE003694	AE003694 Drosophil
31	68.2	7.1	301068	1	AE017231	AE017231 Mycobacte
c 32	59.4	6.2	110000	8	CR382132_38	Continuation (39 o
c 33	56.6	5.9	3014	6	CQ579386	CQ579386 Sequence
c 34	56.6	5.9	3349	2	AC020400	AC020400 Drosophil
c 35	56.2	5.9	3308	6	CQ612719	CQ612719 Sequence
c 36	56.2	5.9	10400	6	CQ599579	CQ599579 Sequence
c 37	56.2	5.9	10418	6	CQ612704	CQ612704 Sequence
c 38	56.2	5.9	10451	6	CQ598352	CQ598352 Sequence
c 39	56.2	5.9	39207	2	AC020418	AC020418 Drosophil
c 40	54.4	5.7	11874	1	AE012528	AE012528 Xanthomon
c 41	51.6	5.4	299300	1	AP005026	AP005026 Streptomy
c 42	50	5.2	303400	1	AP006578	AP006578 Globobact
c 43	49.8	5.2	2000	6	AX655393	AX655393 Sequence
c 44	49	5.1	110000	1	AE017225_45	Continuation (46 o
c 45	49	5.1	110000	1	AE017225_46	Continuation (47 o

ALIGNMENTS

RESULT 1
AR405970
LOCUS AR405970 960 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 12 from patent US 6630330.
ACCESSION AR405970
VERSION AR405970.1 GI:40155041
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 960)
PORRO,D. and Sauer,M.
TITLE Ascorbic acid production from yeast
JOURNAL Patent: US 6630330-A 12 07-OCT-2003;
FEATURES Location/Qualifiers
source 1..960
/organism="unknown"
/mol_type="genomic DNA"

Query Match	100.0%	Score 960;	DB 6;	Length 960;
Best Local Similarity	100.0%	Pred. No. 1.7e-263;		
Matches 960;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	ATGACGAAAATAGAGCTTCGAGCTTTGGGGAACACAGGGCTTAAGGTTAGCGCCGTTGGT	60	
Db	1	ATGACGAAAATAGAGCTTCGAGCTTTGGGGAACACAGGGCTTAAGGTTAGCGCCGTTGGT	60	
Qy	61	TTTGGTGCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGGAAGATGATGCCGTCGCC	120	
Db	61	TTTGGTGCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGGAAGATGATGCCGTCGCC	120	
Qy	121	ACCGTGGCGGAGGCTTTCCGCTCGGTATCAACTTCTTCGACACCTCCCGTATTATGGA	180	
Db	121	ACCGTGGCGGAGGCTTTCCGCTCGGTATCAACTTCTTCGACACCTCCCGTATTATGGA	180	
Qy	181	GGAACTGTCTGAGAAAATCTCTTGTAAGGGAAGCTTTGCAAGTCCCTAGAAAGT	240	
Db	181	GGAACTGTCTGAGAAAATCTCTTGTAAGGGAAGCTTTGCAAGTCCCTAGAAAGT	240	
Qy	241	GACTACATTTGGGCTACTAAGTGTGGTAGATATAAAGAGGTTTGTATTGATTCAGTCTGAG	300	
Db	241	GACTACATTTGGGCTACTAAGTGTGGTAGATATAAAGAGGTTTGTATTGATTCAGTCTGAG	300	

Db 241 GACTACATTGGCTACTAAGTGTGGTAGATATAAAGAGGTTTGGATTTCAGTCTGAG 300
QY 301 AGAGTAAGAAAGAGTATTGACGAGAGCTTTGGAGAGCTTCAGCTTGAATTATGTTGACATA 360
Db 301 AGAGTAAGAAAGAGTATTGACGAGAGCTTTGGAGAGCTTCAGCTTGAATTATGTTGACATA 360
QY 361 CTTGATTTGTCATGACATTGAGTTCCGGTCTCTTGATCAGATGTCGATGGAACAATTCCT 420
Db 361 CTTGATTTGTCATGACATTGAGTTCCGGTCTCTTGATCAGATGTCGATGGAACAATTCCT 420
QY 421 GCTCTTCAGAACTGAAACAAGAGGGAAGACCCGTTTCATTGGTATCACTGGTCTCCG 480
Db 421 GCTCTTCAGAACTGAAACAAGAGGGAAGACCCGTTTCATTGGTATCACTGGTCTCCG 480
QY 481 TTGATATATTTTCACTATATGTTCTTGATCGAGTGCCTCCAGGAGCTTCGATGATATG 540
Db 481 TTGATATATTTTCACTATATGTTCTTGATCGAGTGCCTCCAGGAGCTTCGATGATATG 540
QY 541 TCATATCTGTCATTCAGGGCTTAATGATTCGACGTTGCTGGATTACTACCTTACTTTGAAG 600
Db 541 TCATATCTGTCATTCAGGGCTTAATGATTCGACGTTGCTGGATTACTACCTTACTTTGAAG 600
QY 601 AGCAAGGTGTCGGTGTGATAGTCTTCTCCATTAGCAATGGGCTCCTTACAGAACAA 660
Db 601 AGCAAGGTGTCGGTGTGATAGTCTTCTCCATTAGCAATGGGCTCCTTACAGAACAA 660
QY 661 GGTCTCTCTGAAATGGCACCCTCTCTCCCTGAGCTCAAGTCTGCAAGCAAAAGCCGAGTT 720
Db 661 GGTCTCTCTGAAATGGCACCCTCTCTCCCTGAGCTCAAGTCTGCAAGCAAAAGCCGAGTT 720
QY 721 GCTCACTGCAATCAAAAGGCAAGAGATCAAAAGTTAGCTCTGCAATACAGTTTAGCA 780
Db 721 GCTCACTGCAATCAAAAGGCAAGAGATCAAAAGTTAGCTCTGCAATACAGTTTAGCA 780
QY 781 ACAAGGAGATTTCGTGGTGTGGTGGATGAGCTCTGCTCACAGTAGAAGAAAT 840
Db 781 ACAAGGAGATTTCGTGGTGTGGTGGATGAGCTCTGCTCACAGTAGAAGAAAT 840
QY 841 GTTGACGAGTTTACAGAGCTTGAAAGTCTGGGATGGATCAAGAACTCTGTCTGAGTT 900
Db 841 GTTGACGAGTTTACAGAGCTTGAAAGTCTGGGATGGATCAAGAACTCTGTCTGAGTT 900
QY 901 GAAGCTATTCGAGCTGTAAGAAATCTGACATGCGCAAGTGGAAATCCATCAGAACTAA 960
Db 901 GAAGCTATTCGAGCTGTAAGAAATCTGACATGCGCAAGTGGAAATCCATCAGAACTAA 960

RESULT 2
LOCUS AX259907 960 bp DNA linear PAT 26-OCT-2001
DEFINITION Sequence 4 from Patent WO0172974.
ACCESSION AX259907
VERSION AX259907.1 GI:16508970
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1
REFERENCE 1
AUTHORS Smirnov, N.S. and Wheeler, G.L.
TITLE L-galactose dehydrogenase from Arabidopsis thaliana, and uses thereof
JOURNAL Patent: WO 0172974-A 4 04-OCT-2001;
Ascorbex Limited (GB)
FEATURES
source
1..960 Location/Qualifiers
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
1..960
/note="unnamed protein product"
/codon_start=1
CDS

/protein_id="CAD10214.1"
/db_xref="GI:16508971"
/translation="MTKIELRALGNTGLKVSAVFGASPLGVSFPGVAEDDAVATVRE
AFRLGINFDTSPYGGTGLSEKMLGKGLKALQVPRSDYIVATKGRYKGFDFSAERV
RKSIDESLRQLQDVIDILHCHDIEFGVSDQIVSETIPALOKLQKQKTRIGTGLP
LDJFTYLDRLVPRGTVDVILSYCHVGNLTDLLPVLKSGVGVISASPLAMGLLT
EQGPPEHPASPPELKSAAVAHCKSKKKITKLALQYSLANKEISSVLVGMSSVSQ
VEENVAATVELESIGMQDETUSEVEALDEPVKNLTFWFSGHQN"
ORIGIN
Query Match 100.0%; Score 960; DB 6; Length 960;
Best Local Similarity 100.0%; Pred. No. 1.7e-263;
Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGACGAAAATAGAGCTTCGAGCTTTGGGGAACAACAGGGCTTAAAGTTAGCGCGCTTGGT 60
Db 1 ATGACGAAAATAGAGCTTCGAGCTTTGGGGAACAACAGGGCTTAAAGTTAGCGCGCTTGGT 60
QY 61 TTTGGTGGCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCCGCGAAGATGATGCCGTCGCC 120
Db 61 TTTGGTGGCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCCGCGAAGATGATGCCGTCGCC 120
QY 121 ACCGTGCGGAGGCTTTCCGCTCGGTATCAACTTCTTCGACACCTCCCGCTATTATGGA 180
Db 121 ACCGTGCGGAGGCTTTCCGCTCGGTATCAACTTCTTCGACACCTCCCGCTATTATGGA 180
QY 181 GGAACACCTGTCTGAGAAAATGCTTTGGTAAAGGCACTAAAGGCTTTGCAAGTCCCTAGAAAT 240
Db 181 GGAACACCTGTCTGAGAAAATGCTTTGGTAAAGGCACTAAAGGCTTTGCAAGTCCCTAGAAAT 240
QY 241 GACTACATTGGCTTACTAAGTGTGGTAGATATAAAGAGGTTTGGATTTCAGTGTCTGAG 300
Db 241 GACTACATTGGCTTACTAAGTGTGGTAGATATAAAGAGGTTTGGATTTCAGTGTCTGAG 300
QY 301 AGAGTAAGAAAGAGTATTGACGAGAGCTTTGAGAGGCTTCAGCTTGAATTATGTTGACATA 360
Db 301 AGAGTAAGAAAGAGTATTGACGAGAGCTTTGAGAGGCTTCAGCTTGAATTATGTTGACATA 360
QY 361 CTTCAATTGCCATGACATTGAGTTCCGGTCTCTTGATCAGATGTCGATGGAACAATTCCT 420
Db 361 CTTCAATTGCCATGACATTGAGTTCCGGTCTCTTGATCAGATGTCGATGGAACAATTCCT 420
QY 421 GCTCTTCAGAACTGAAACAAGAGGGAAGACCCGTTTCATTGGTATCACTGGTCTTCG 480
Db 421 GCTCTTCAGAACTGAAACAAGAGGGAAGACCCGTTTCATTGGTATCACTGGTCTTCG 480
QY 481 TTGATATATTTTCACTATATGTTCTTGATCGAGTGCCTCCAGGAGCTTCGATGATATG 540
Db 481 TTGATATATTTTCACTATATGTTCTTGATCGAGTGCCTCCAGGAGCTTCGATGATATG 540
QY 541 TCATATCTGTCATTCAGGGCTTAATGATTCGACGTTGCTGGATTACTACCTTACTTTGAAG 600
Db 541 TCATATCTGTCATTCAGGGCTTAATGATTCGACGTTGCTGGATTACTACCTTACTTTGAAG 600
QY 601 AGCAAGGTGTCGGTGTGATAGTCTTCTCCCTGAGCTCAAGTCTGCAAGCAAAAGCCGAGTT 720
Db 601 AGCAAGGTGTCGGTGTGATAGTCTTCTCCCTGAGCTCAAGTCTGCAAGCAAAAGCCGAGTT 720
QY 721 GCTCACTGCAAAATCAAAAGGCAAGAGATCAAAAGTTAGCTCTGCAATACAGTTTAGCA 780
Db 721 GCTCACTGCAAAATCAAAAGGCAAGAGATCAAAAGTTAGCTCTGCAATACAGTTTAGCA 780
QY 781 AACAGGAGATTTCGTGGTGTGGTGGATGAGCTCTGCTCACAGTAGAAGAAAT 840
Db 781 AACAGGAGATTTCGTGGTGTGGTGGATGAGCTCTGCTCACAGTAGAAGAAAT 840
QY 841 GTTGACGAGTTTACAGAGCTTGAAAGTCTGGGATGGATCAAGAACTCTGTCTGAGTT 900
Db 841 GTTGACGAGTTTACAGAGCTTGAAAGTCTGGGATGGATCAAGAACTCTGTCTGAGTT 900


```
Qy 901 GAAGCTATTCTCGAGCCTGTAAAGAACTCTGACATGCGCAAGTGAATCCATCAGAACTAA 960
Db 901 GAAGCTATTCTCGAGCCTGTAAAGAACTCTGACATGCGCAAGTGAATCCATCAGAACTAA 960

RESULT 3
AX374959
LOCUS AX374959 960 bp DNA linear PAT 01-MAR-2002
DEFINITION Sequence 12 from Patent WO0210425.
ACCESSION AX374959
VERSION AX374959.1 GI:19169805
KEYWORDS
SOURCE
ORGANISM Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
REFERENCE
AUTHORS Porro,D. and Sauer,M.
TITLE Ascorbic acid production from yeasts
JOURNAL Patent: WO 0210425-A 12 07-FEB-2002;
Biopolo S.C.A.R.L. (IT)
FEATURES
source
1..960
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"

ORIGIN
Query Match 100.0%; Score 960; DB 6; Length 960;
Best Local Similarity 100.0%; Pred. No. 1.7e-263;
Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCGAAATAGAGCTTCGAGCTTGGGAAACACAGGCTTAAGTTAGCGCGTGGT 60
Db 1 ATGCGAAATAGAGCTTCGAGCTTGGGAAACACAGGCTTAAGTTAGCGCGTGGT 60

Qy 61 TTGTGCTCTCCGCTCGGAGTGTCTTCGGTCCAGTCGCGAAGATGATGCGGTGCC 120
Db 61 TTGTGCTCTCCGCTCGGAGTGTCTTCGGTCCAGTCGCGAAGATGATGCGGTGCC 120

Qy 121 ACCGTGCGGAGGCTTCCGCTCGGTATCAACTTCTTCGACACCTCCCGTATTATGGA 180
Db 121 ACCGTGCGGAGGCTTCCGCTCGGTATCAACTTCTTCGACACCTCCCGTATTATGGA 180

Qy 181 GGAACACTCTGAGAAATGCTTGGTAAGGACTTAAGGCTTTCGAGTCCCTAGAGT 240
Db 181 GGAACACTCTGAGAAATGCTTGGTAAGGACTTAAGGCTTTCGAGTCCCTAGAGT 240

Qy 241 GACTACATTGGCTACTAAGTGTGGTAGATATAAAGAGGTTTGTATTCAGTGCCTGAG 300
Db 241 GACTACATTGGCTACTAAGTGTGGTAGATATAAAGAGGTTTGTATTCAGTGCCTGAG 300

Qy 301 AGAGTAAGAAAGATATTGACGAGCTTGAGAGGCTTCAGTGTGATATTATGTCACATA 360
Db 301 AGAGTAAGAAAGATATTGACGAGCTTGAGAGGCTTCAGTGTGATATTATGTCACATA 360

Qy 361 CTTGATGCCATGACATTGAGTTCGGTCTCTTGTATCAGATGTGAGTGAACAATTCCT 420
Db 361 CTTGATGCCATGACATTGAGTTCGGTCTCTTGTATCAGATGTGAGTGAACAATTCCT 420

Qy 421 GCTCTTCAGAACTGAAACAAGAGGGGAGACCGGTTTCATTGGTATCAGTTCCTCCG 480
Db 421 GCTCTTCAGAACTGAAACAAGAGGGGAGACCGGTTTCATTGGTATCAGTTCCTCCG 480

Qy 481 TTAGATATTTTCACTATGTTCTTGATCGAGTGCCTCCAGGACTGTCGATGTGATATTG 540
Db 481 TTAGATATTTTCACTATGTTCTTGATCGAGTGCCTCCAGGACTGTCGATGTGATATTG 540

Qy 541 TCATCTCTGATACCGGCTTAAGTATCGAGTTCGCTGATTTTACTACCTTACTTGAAG 600
Db 541 TCATCTCTGATACCGGCTTAAGTATCGAGTTCGCTGATTTTACTACCTTACTTGAAG 600
```

```
Qy 601 AGCAAAGGTGTGGGTGTGATAAGTCTTCTCCATTAGCAATGGCCCTCTTACAGAACAA 660
Db 601 AGCAAAGGTGTGGGTGTGATAAGTCTTCTCCATTAGCAATGGCCCTCTTACAGAACAA 660

Qy 661 GGTCTCTCTGAATGGGCACTCTTCCCTGAGCTCAAGTCTGCAAGCAAGCCGAGTT 720
Db 661 GGTCTCTCTGAATGGGCACTCTTCCCTGAGCTCAAGTCTGCAAGCAAGCCGAGTT 720

Qy 721 GCTCACTGCAAAATCAAAAGGCGCAAGATCAAAAGTTAGCTCTGCAATCAGTTTAGCA 780
Db 721 GCTCACTGCAAAATCAAAAGGCGCAAGATCAAAAGTTAGCTCTGCAATCAGTTTAGCA 780

Qy 781 AACAAAGGAGATTTCTCGTGGTGTGGATGAGCTCTCTCACAAGTAGAAGAAAT 840
Db 781 AACAAAGGAGATTTCTCGTGGTGTGGATGAGCTCTCTCACAAGTAGAAGAAAT 840

Qy 841 GTTCACAGCTTACAGAGCTTGAAGTCTGGGATGATCAAGAACTCTCTCTGAGGTT 900
Db 841 GTTCACAGCTTACAGAGCTTGAAGTCTGGGATGATCAAGAACTCTCTCTGAGGTT 900

Qy 901 GAAGCTATTCTCGAGCCTGTAAAGAAATCTGACATGCGCAAGTGAATCCATCAGAACTAA 960
Db 901 GAAGCTATTCTCGAGCCTGTAAAGAAATCTGACATGCGCAAGTGAATCCATCAGAACTAA 960

RESULT 4
ATH417563 960 bp mRNA linear PLN 26-JUL-2002
LOCUS Arabidopsis thaliana mRNA for L-galactose dehydrogenase (L-GalDH
DEFINITION gene).
ACCESSION AJ417563
VERSION AJ417563.1 GI:16555789
KEYWORDS L-galactose dehydrogenase; L-GalDH gene.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
Gatzek, S., Wheeler, G.L. and Smirnov, N.
Antisense suppression of l-galactose dehydrogenase in Arabidopsis
thaliana provides evidence for its role in ascorbate synthesis and
reveals light modulated l-galactose synthesis
Plant J. 30 (5), 541-553 (2002)
JOURNAL
MEDLINE 22043773
PUBMED 12047629
REFERENCE
2 (bases 1 to 960)
AUTHORS Gatzek, S.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-2001) Gatzek S., School of Biological Sciences,
University of Exeter, Exeter, EX4 4PS, UNITED KINGDOM
FEATURES
source
1..960
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
1..960
/gene="L-GalDH"
1..960
/gene="L-GalDH"
/functions=oxidises L-galactose at C1, forming
L-galactono-1,4-lactone"
/codon_start=1
/evidence=experimental
/product="L-galactose dehydrogenase"
/protein_id="CAD10386.1"
/db_xref="GI:16555790"
/db_xref="UniProt/TcEMBL:O81884"
/tranlation="MTKLELRALGNTGLKVSAGVFGASPLGSVFGVAEDDAVATVRE
AFRLGNFTPTSPYGGTSLSEKMLKGLKALQVPRSDYIIVATKGRYKEGDFSAERV
RKSIDESLRQLQDYLVDLHCHDIEFGLDQIVSETIPALQKLKQEGKTRFTIGLPL
LDIFTYVLDVRVPPGTVDVILSYCHYGVNDSTLDDLPLYLKSGVGVISASPLWGLLT
```

BOGPPEWHPASPESKASKAAVAHCKSKGKITKLALQYSLANKBEISSVLVGMSSVSQ
VEENVAAVTELESIGMDQETLSEVAILEPKNLTPWPSGIHQN"

ORIGIN	Query Match Best Local Similarity Matches 960; Conservative 0; Mismatches 0; Length 960; Indels 0; Gaps 0;
QY	1 ATGACGAAATAGAGCTTCGAGCTTTGGGAAACACAGGCTTAAGCTTAGCGCCGCTTGGT 60
Db	1 ATGACGAAATAGAGCTTCGAGCTTTGGGAAACACAGGCTTAAGCTTAGCGCCGCTTGGT 60
QY	61 TTTGGTGCCCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCCGGAAGATGATGCCGCTCGCC 120
Db	61 TTTGGTGCCCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCCGGAAGATGATGCCGCTCGCC 120
QY	121 ACCGTGCGGAGGCTTCGGTCCGATCACTTCCTTCGACACCTCCCGTATTATATGGA 180
Db	121 ACCGTGCGGAGGCTTCGGTCCGATCACTTCCTTCGACACCTCCCGTATTATATGGA 180
QY	181 GGAAACACTGCTCGAGAAATGCTTCGTAAGGACTAAAGGCTTTGCAAGTCCCTAGAAAGT 240
Db	181 GGAAACACTGCTCGAGAAATGCTTCGTAAGGACTAAAGGCTTTGCAAGTCCCTAGAAAGT 240
QY	241 GACTACATTGTGGCTACTAAGTGTGGTAGATATAAAGAGGTTTGTATTCAGTCTGAG 300
Db	241 GACTACATTGTGGCTACTAAGTGTGGTAGATATAAAGAGGTTTGTATTCAGTCTGAG 300
QY	301 AGAGTAAGAAAGATGATGACAGAGCTTGGAGGCTTCAGCTGATGATGTTGACATA 360
Db	301 AGAGTAAGAAAGATGATGACAGAGCTTGGAGGCTTCAGCTGATGATGTTGACATA 360
QY	361 CTTTCATTGTCATGACATTGAGTTCGGGCTCTTTGATCAGATTGTGAGTGAAACAATTCCT 420
Db	361 CTTTCATTGTCATGACATTGAGTTCGGGCTCTTTGATCAGATTGTGAGTGAAACAATTCCT 420
QY	421 GCTCTTCAGAAACTGAACAGAGGGAAGACCCGGTTCATTGGTATCACTGGTCTTCGG 480
Db	421 GCTCTTCAGAAACTGAACAGAGGGAAGACCCGGTTCATTGGTATCACTGGTCTTCGG 480
QY	481 TTAGATATTTTCACTATGTTCTTCATCGAGTGCTCCAGGACGTGCGATGTGATTTG 540
Db	481 TTAGATATTTTCACTATGTTCTTCATCGAGTGCTCCAGGACGTGCGATGTGATTTG 540
QY	541 TCATACATGTCATTAAGGCTTAATGATTCGAGCTTGCTGGATTTACTACTTACTTTGAAG 600
Db	541 TCATACATGTCATTAAGGCTTAATGATTCGAGCTTGCTGGATTTACTACTTACTTTGAAG 600
QY	601 AGCAAAGGTGGGTGTGATAGTGTCTTCATAGCATGGGCTCTCTACAGAACAA 660
Db	601 AGCAAAGGTGGGTGTGATAGTGTCTTCATAGCATGGGCTCTCTACAGAACAA 660
QY	661 GGTCTCTGTAATGGCACCCTCTTCCCTGAGCTCAAGTCTGCAAGCAAGCCGAGTT 720
Db	661 GGTCTCTGTAATGGCACCCTCTTCCCTGAGCTCAAGTCTGCAAGCAAGCCGAGTT 720
QY	721 GCTCACTGCAATCAAGGCGCAAGAGATCAAAAAGTTAGCTCTGCAATACAGTTTAGCA 780
Db	721 GCTCACTGCAATCAAGGCGCAAGAGATCAAAAAGTTAGCTCTGCAATACAGTTTAGCA 780
QY	781 AACAGAGGATTTTGGTGGGTGGATGAGCTCTGCTCACAGGTAGAGAAAT 840
Db	781 AACAGAGGATTTTGGTGGGTGGATGAGCTCTGCTCACAGGTAGAGAAAT 840
QY	841 GTTGACAGGATTTACAGAGCTTCAAGTCTGGGATGGATCAAGAACTCTGCTCAGGTT 900
Db	841 GTTGACAGGATTTACAGAGCTTCAAGTCTGGGATGGATCAAGAACTCTGCTCAGGTT 900
QY	901 GAAGCTATTCTCGAGCTGTAAGAATCTGACATGCCCAAGTGGATCCATCAGAACTAA 960
Db	901 GAAGCTATTCTCGAGCTGTAAGAATCTGACATGCCCAAGTGGATCCATCAGAACTAA 960

RESULT 5
AY090337
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AY090337
Arabidopsis thaliana AT4g33670/Ti6L1_160 mRNA linear PLN 24-MAR-2002
AY090337
AY090337.1 GI:19699263
FLI_CDNA.
Arabidopsis thaliana (thale cress)

Arabidopsis thaliana
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopses.

1 (bases 1 to 960)

Chen, R., Chen, H., Kim, C.J., Meyers, M.C., Shinn, P., Banh, J.,
Bowser, L., Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D.,
Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G.,
Kawai, J., Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M.,
Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T.,
Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C.,
Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W.,
Theologis, A. and Ecker, J.R.

Arabidopsis ORF clones

Unpublished

2 (bases 1 to 960)

Chen, R., Chen, H., Kim, C.J., Meyers, M.C., Shinn, P., Banh, J.,
Bowser, L., Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D.,
Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G.,
Kawai, J., Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M.,
Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T.,
Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C.,
Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W.,
Theologis, A. and Ecker, J.R.

Direct Submission

Submitted (14-MAR-2002) Salk Institute Genomic Analysis Laboratory
(SIGNAL), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA

RIKEN Genomic Sciences Center (GSC) Consortium members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
Arabidopsis Full-Length cDNA') : Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGSC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Chen, R., Chen, H.,
Kim, C.J., Meyers, M.C., Shinn, P., Banh, J., Bowser, L., Chan, M.M.,
Chang, E., Dale, J.M., Deng, J.M., Goldsmith, A.D., Jones, T.,
Karlin-Neumann, G., Lam, B., Lee, J.M., Lin, J., Miranda, M., Nguyen, M.,
Onodera, C.S., Palm, C.J., Quach, H.L., Southwick, A., Tang, C.C.,
Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Davis, R.W.,
Theologis, A., and Ecker, J.R.

Chen, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to
this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)
contributed equally to this work as PIs.

Location/Qualifiers

1. .960
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="4"
/clone="U13146"
/ecotype="Columbia"
/note="This clone is in pUNI 51"

1. .960
/note="unknown protein"

/codon_start=1
/product="AT4g33670/Ti6L1_160"
/protein_id="AAL90998.1"
/db_xref="GI:19699264"
/translation="MTKIELRALNGTLKVSAGVFGASPLGSVFGPVAEDDAVATVRE
AFRLGNPFDTSPPYVGGTSLSEKMLGKLGKALQVPRSDYIVATKCRVYEGDFSAERV
RKSDIESLERLQDLYDVLHCHDIEFGSLDQIVSETIPALQKLQEGKTRFIGITGLP

LDIIFYVLDRVPPGTVDVILSYCHYGVNDSTLLDLLPYLKSXGVGVISASPLAMGLLT
BOGPPEHPASPILKSAKAAVAHCKSKKTKTIALQYSLANKEISSVLVGMSSVSQ
VEENVAATELSLGMDOETLSEVAILEPVLKNTWPSGIHQN"

ORIGIN	Query Match		Best Local Similarity		Matches 960; Conservative		0; Mismatches		0; Indels		0; Gaps		0;	
	100.0%; Score 960; DB 8; Length 960;		100.0%; Pred. No. 1.7e-263;		0;		0;		0;		0;		0;	
Qy	1	ATGACGAAATAGAGCTTCGAGCTTTGGGGAACACAGGCTTAAAGTTAGCGCCGTTGGT	60											
Db	1	ATGACGAAATAGAGCTTCGAGCTTTGGGGAACACAGGCTTAAAGTTAGCGCCGTTGGT	60											
Qy	61	TTTGGTGCTCTCCGCTCGGAAGTCTTCGGTCCAGTCGCGGAAGATGATGCCGTCGCC	120											
Db	61	TTTGGTGCTCTCCGCTCGGAAGTCTTCGGTCCAGTCGCGGAAGATGATGCCGTCGCC	120											
Qy	121	ACCGTGCGGAGGCTTCCGCTCGGTATCACTTCTTCGCACACCTCCCGTATTATGGA	180											
Db	121	ACCGTGCGGAGGCTTCCGCTCGGTATCACTTCTTCGCACACCTCCCGTATTATGGA	180											
Qy	181	GGAACACTGTCTGAGAAATGCTTCGTAAGGACTTAAAGGCTTTGCAAGTCCCTAGAAGT	240											
Db	181	GGAACACTGTCTGAGAAATGCTTCGTAAGGACTTAAAGGCTTTGCAAGTCCCTAGAAGT	240											
Qy	241	GACTACATTTGGCTACTAAGTGTGGTAGATATAAAGAGGTTTGTATTCAGTCTGAG	300											
Db	241	GACTACATTTGGCTACTAAGTGTGGTAGATATAAAGAGGTTTGTATTCAGTCTGAG	300											
Qy	301	AGAGTAAGAGAGTATTGACGAGAGCTTGGAGGCTTCAGCTTGATATGTTGACATA	360											
Db	301	AGAGTAAGAGAGTATTGACGAGAGCTTGGAGGCTTCAGCTTGATATGTTGACATA	360											
Qy	361	CTTCATTCGCATGACATTTGGTTCGGTCTCTTGATCAGATGTGTCAGTGAACAACTTCT	420											
Db	361	CTTCATTCGCATGACATTTGGTTCGGTCTCTTGATCAGATGTGTCAGTGAACAACTTCT	420											
Qy	421	GCTCTTCAGAACTCAAAACAGAGGGGAGACCCGGTTTCATTTGATCAGTCTGCTCCG	480											
Db	421	GCTCTTCAGAACTCAAAACAGAGGGGAGACCCGGTTTCATTTGATCAGTCTGCTCCG	480											
Qy	481	TTAGATATTTTCACTTATGTTTGAATGCTTTCATTTGATCAGTCTGCTGATGATTTG	540											
Db	481	TTAGATATTTTCACTTATGTTTGAATGCTTTCATTTGATCAGTCTGCTGATGATTTG	540											
Qy	541	TCATACCTGTCATTCAGGCTTAAATGATTCGAGTTGCTGATTTACTACCTTACTTTGAAG	600											
Db	541	TCATACCTGTCATTCAGGCTTAAATGATTCGAGTTGCTGATTTACTACCTTACTTTGAAG	600											
Qy	601	AGCAAGGTGTGGTGTGATAGTCTTCTCCATTTAGCAATGGGCTCTCTTACAGAACAA	660											
Db	601	AGCAAGGTGTGGTGTGATAGTCTTCTCCATTTAGCAATGGGCTCTCTTACAGAACAA	660											
Qy	661	GGTCTCTCTGATGCGACCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAGCCGAGTT	720											
Db	661	GGTCTCTCTGATGCGACCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAGCCGAGTT	720											
Qy	721	GCTCACTGCAATCAAGGGCAAGAGATCAAAAGTTAGCTCTGCAATACAGTTTACGA	780											
Db	721	GCTCACTGCAATCAAGGGCAAGAGATCAAAAGTTAGCTCTGCAATACAGTTTACGA	780											
Qy	781	AACAGGAGATTTCTCGGTGTGTTGGTGTGGATGAGCTCTGTCTCA CAGGTAGAAGAAAT	840											
Db	781	AACAGGAGATTTCTCGGTGTGTTGGTGTGGATGAGCTCTGTCTCA CAGGTAGAAGAAAT	840											
Qy	841	GTTCAGAGATTTACAGAGCTTGAAGCTGGGATGGATCAAGAACTCTGTCTGAGTT	900											
Db	841	GTTCAGAGATTTACAGAGCTTGAAGCTGGGATGGATCAAGAACTCTGTCTGAGTT	900											
Qy	901	GAAGCTATTCTGAGCCTGTAAGAACTCTGACATGCCCAAGTGGAAATCCATCAGAACTAA	960											
Db	901	GAAGCTATTCTGAGCCTGTAAGAACTCTGACATGCCCAAGTGGAAATCCATCAGAACTAA	960											

RESULT 6
AY050377

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

FEATURES
source

5'UTR
CDS

AY050377
Arabidopsis thaliana AT4g33670/T16L1_160 mRNA, complete cds.
AY050377
AY050377.1 GI:15215697
FLI CDNA.
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1092)
Koesema, E., Chen, H., Cheuk, R., Kim, C.J., Meyers, M.C., Shinn, P.,
Ban, J., Bowser, L., Carninci, P., Dale, J.M., Goldsmith, A.D.,
Hayashizaki, Y., Ishida, J., Jiang, P.X., Jones, T., Kamiya, A.,
Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X.,
Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J.,
Pham, P.K., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A.,
Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S.,
Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.
Arabidopsis c NA clones
Unpublished
2 (bases 1 to 1092)
Koesema, E., Chen, H., Cheuk, R., Kim, C.J., Meyers, M.C., Shinn, P.,
Ban, J., Bowser, L., Carninci, P., Dale, J.M., Goldsmith, A.D.,
Hayashizaki, Y., Ishida, J., Jiang, P.X., Jones, T., Kamiya, A.,
Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X.,
Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J.,
Pham, P.K., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A.,
Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S.,
Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.
Direct Submission
Submitted (06-AUG-2001) Salk Institute Genomic Analysis Laboratory
(SIGNAL), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA

RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
Arabidopsis Full-length cDNA') : Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGSC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Koesema, E., Chen, H.,
Cheuk, R., Kim, C.J., Meyers, M.C., Shinn, P., Ban, J., Bowser, L.,
Dale, J.M., Goldsmith, A.D., Jiang, P.X., Jones, T., Karlin-Neumann, G.,
Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Nguyen, M.,
Onodera, C.S., Palm, C.J., Pham, P.K., Quach, H.L., Southwick, A.,
Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S.,
Davis, R.W., Theologis, A., and Ecker, J.R.

Koesema, E. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally
to this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)
contributed equally to this work as PIs.

Location/Qualifiers
1. .1092
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="4"
/clones="RAFL07-17-C15(R13146)"
/ecotype="Columbia"
1. .7
8. .967
/note="unknown protein"
/codon_start=1
/product="AT4G33670/T16L1_160"
/protein_id="AAK91395.1"
/db_xref="GI:15215698"

translation="MTKIELRALNGTGLKVSAGVGFSGPLSGFVGFVADDAVTVRE
AFRLGINFFDTSPPYGGTSLSEKMLGKALQVPRSDYIVATKGRYKEGDFSAERV

RKSIDESLERLQLDVVDILHCHDIEFGSLDQIVSETIPALQKLQKQKTRFIFGTGLP LDIFYVLDVRVPPGTVVILSYCHYGVNDSTLLDLLPYLKSQKGVIVSASPLAMGLLT EQGPEHPAPASKAAVAHCKSKGKITKLQYSLANKEISSVLVGMSSVSQ VEENVAAVTELESLEMDQETLSEVAILEPKNLTPWPGIHQN" 968. .1092									
3'UTR									
ORIGIN									
Query Match 100.0%; Score 960; DB 8; Length 1092; Best Local Similarity 100.0%; Pred. No. 1.7e-263; Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	ATGACGAAAATAGAGCTTCGAGCTTTGGGAACACAGGGCTTAAGGTTAGCGCCGTTGGT	60						
Db	8	ATGACGAAAATAGAGCTTCGAGCTTTGGGGAACACAGGGCTTAAGGTTAGCGCCGTTGGT	67						
Qy	61	TTTGGTGCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGCAAGATGATGCCGTCGCC	120						
Db	68	TTTGGTGCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGCAAGATGATGCCGTCGCC	127						
Qy	121	ACCGTCGCGAGGCTTCCGCTCGGTATCAACTCTTTGACACACCTCCCGTATTATGGA	180						
Db	128	ACCGTCGCGAGGCTTCCGCTCGGTATCAACTCTTTGACACACCTCCCGTATTATGGA	187						
Qy	181	GGAACTACTGTCGAGAAAATGCTTGGTAAGGACTAAAGGCTTTCGAAGTCCCTAGAGT	240						
Db	188	GGAACTACTGTCGAGAAAATGCTTGGTAAGGACTAAAGGCTTTCGAAGTCCCTAGAGT	247						
Qy	241	GACTACATTTGCGTACTAAGTGTGTAGATATAAGAAAGGTTTGTATTCAGTCTGAG	300						
Db	248	GACTACATTTGCGTACTAAGTGTGTAGATATAAGAAAGGTTTGTATTCAGTCTGAG	307						
Qy	301	AGAGTAAGAAAGATATTGACGAGAGCTTTGGAGAGCTTCAGCTTGATATTGTCACATA	360						
Db	308	AGAGTAAGAAAGATATTGACGAGAGCTTTGGAGAGCTTCAGCTTGATATTGTCACATA	367						
Qy	361	CTTCATTTGCCATGACATTTAGTTCGGTCTCTTGATCAGATCTGAGTGAACAATTCCT	420						
Db	368	CTTCATTTGCCATGACATTTAGTTCGGTCTCTTGATCAGATCTGAGTGAACAATTCCT	427						
Qy	421	GCTCTTCAGAACTGAAACAGAGGGGAAGACCGGTTTCATTTGGTATCACTGGTCTTCG	480						
Db	428	GCTCTTCAGAACTGAAACAGAGGGGAAGACCGGTTTCATTTGGTATCACTGGTCTTCG	487						
Qy	481	TTAGATATTTTCACTATGTTCTTGATCGAGTGCTTCCAGGACCTGTCGATGTATATTG	540						
Db	488	TTAGATATTTTCACTATGTTCTTGATCGAGTGCTTCCAGGACCTGTCGATGTATATTG	547						
Qy	541	TCATCTGTCATTAAGCGGTTAATGATTCGAGTTGCTGGATTTACTACCTTACTTTGAAG	600						
Db	548	TCATCTGTCATTAAGCGGTTAATGATTCGAGTTGCTGGATTTACTACCTTACTTTGAAG	607						
Qy	601	AGCAAGAGTGTGGGTGTGATAGTGTCTTCCATTAGCAATGGGCTCTCTTACAGAACAA	660						
Db	608	AGCAAGAGTGTGGGTGTGATAGTGTCTTCCATTAGCAATGGGCTCTCTTACAGAACAA	667						
Qy	661	GGTCTCTTGAATGGACCCCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAGCCGAGTT	720						
Db	668	GGTCTCTTGAATGGACCCCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAGCCGAGTT	727						
Qy	721	GCTCACTGCAATCAAGGGCGAAGATGACAAAGTTAGCTCTGCAATACAGTTTATGCA	780						
Db	728	GCTCACTGCAATCAAGGGCGAAGATGACAAAGTTAGCTCTGCAATACAGTTTATGCA	787						
Qy	781	AACAAGGATTTTCGTCGGTGTGGTTGGGATGAGCTCTGTCTCAAGGTAGAGAAAT	840						
Db	788	AACAAGGATTTTCGTCGGTGTGGTTGGGATGAGCTCTGTCTCAAGGTAGAGAAAT	847						
Qy	841	GTTGACGAGTTTACAGAGCTTGAAAGTCTGGGATGGATCAAGAACTCTGTCTGAGGTT	900						
Db	848	GTTGACGAGTTTACAGAGCTTGAAAGTCTGGGATGGATCAAGAACTCTGTCTGAGGTT	907						
Qy	901	GAAGCTATTTCTGAGCCTCTAAGATCTGACATGCCAAGTGGATCCATCAGAACTAA	960						

Db	908	GAAGCTATTTCTGAGCCTGTAAAGATCTGACATGGCCCAAGTGGNATCCATCAGAACTAA	967						
RESULT 7									
LOCUS									
AV176585 1215 bp mRNA linear PLN 13-JAN-2003									
DEFINITION									
Actinidia deliciosa galactose dehydrogenase mRNA, complete cds.									
ACCESSION									
AV176585									
VERSION									
AV176585.1 GI:27728730									
KEYWORDS									
SOURCE									
ORGANISM									
Actinidia deliciosa									
Actinidia deliciosa									
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; Ericales; Actinidiaceae; Actinidia.									
REFERENCE									
AUTHORS									
1 (bases 1 to 1215) Beuning,L., Bowen,J., Crowhurst,R., Gleave,A., MacRae,E., Newcomb,R., Perera,S., Ross,G., Snowden,K., Walton,E. and Yauk,Y.K.									
TITLE									
A gene database from Fruit Tree Species									
JOURNAL									
REFERENCE									
AUTHORS									
2 (bases 1 to 1215) Laiing,W.A. and MacRae,E.									
TITLE									
Direct Submission									
JOURNAL									
REFERENCE									
AUTHORS									
3 (bases 1 to 1215) Unpublished Kiwifruit Leaf Galactose Dehydrogenase									
TITLE									
Submitted (08-NOV-2002) Postharvest and Food Science, The Horticulture and Food Research Institute of New Zealand Ltd, PB 92169, Auckland, New Zealand									
FEATURES									
source									
1..1215									
/organism="Actinidia deliciosa"									
/mol_type="mRNA"									
/cultivar="Hayward"									
/db_xref="taxon:3627"									
/tiissue_type="developing shoot buds from extending water buds"									
32..991									
/function="oxidizes L-galactose at C1 to L-galactono-1,4-lactone"									
/codon_start=1									
/evidence=experimental									
/product="galactose dehydrogenase"									
/protein_id="AAO18639.1"									
/db_xref="GI:27728731"									
/translation="MTTDLRLPLGNTGLKLSVGFSGAPLGNVFGVSDHDAIASVRE ALRLGNPFDTSVYGGTSLSEKVLGKALGVPRNEYIVSTKCRVYAEQFPFSAERV TKSPDESRLQLDYVDILOCHDIEFGSLDOI VNETIPALOKLKEAGKIRFTIGTGLP LGVFTYVLDVRVPPGTVVILSYCHYSINDSTLEDLLPYLKSQKGVIVSASPLAMGLLT ESGPPPEHPASPFLKAAACAAAHAHCKEGRNISKIAMQYLSNKKDISSILVGMNSVKQ VEENVAANELATFGKDEKTVSEIEEILPKVKNQTLWSGIQQI"									
ORIGIN									
Query Match 60.7%; Score 582.4; DB 8; Length 1215; Best Local Similarity 75.4%; Pred. No. 2.7e-155; Matches 724; Conservative 0; Mismatches 236; Indels 0; Gaps 0;									
Qy	1	ATGACGAAAATAGAGCTTCGAGCTTTGGGGAACACAGGGCTTTAAGGTTAGCGCCGTTGGT	60						
Db	32	ATGACAAACCTTAGACCTCGACCCCTCGGGAACACGGGCTTGAAGCTCGGTGGGC	91						
Qy	61	TTTGGTGCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGCAAGATGATGCCGTCGCC	120						
Db	92	TTGCGAGCTCTACCTCTGGGCAACGCTCTTTGGCCAGTCTCCGACCAACGACCCATCGCC	151						
Qy	121	ACCGTCGCGAGGCTTTCGCTCGGTATCAACTCTTCGACACCTCCCGTATTATGGA	180						
Db	152	TCGCTCCGCGAGGCCCTTCGCTCGGCATCAACTCTTCGACACCTCTCCGATTATGGA	211						
Qy	181	GGAACTACTGTCGAGAAAATGCTTGGTAAGGACTAAAGGCTTTCGAAGTCCCTAGAGT	240						
Db	212	GGGACATTTGCGAGAAAATGCTTGGGGAAGGCACTAAAAGCTCTGGAGTGGCGGGAAC	271						

Qy	241	GACTACATTTGGCTACTAAGTGTGGTATAGATATAAAGAGGTTTGTGATTTAGTTCAGTCTGAG	300
Db	272	GAGTACATTTGTGCGACCAAGTGTGGAGGATATGACAGAGGCTTTGATTTTTCAGTCCGAG	331
Qy	301	AGAGTAAGAAAGATATTGACAGAGCTTGGAGAGCTTTCAGCTTGTATGATTTGATGACATA	360
Db	332	AGAGTGACTAAGAGCTTTGATGAGAGCTTGGAGAGGTTTTCAGCTCGACTGATTTGATATA	391
Qy	361	CTTCATTCCTCCATGACATTTGGGTCCTTGTATGATGATGATGATGATGATGATGATGATGAT	420
Db	392	TGCAATGCCATGATATTGATTTGGGTCCTTGTATGATGATGATGATGATGATGATGATGATGAT	451
Qy	421	GCTCTTCAGAACTGAAACAGAGAGGAGACCCGGTTTCATTTGGTATCACTGGTCTTCCG	480
Db	452	GCTCTTCAGAACTGAAACAGAGAGGAGACCCGGTTTCATTTGGTATCACTGGTCTTCCG	511
Qy	481	TTAGATATTTTCATTTGCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT	540
Db	512	CTGGAGATTTTACCTATGCTGATCGGTTTCCAGGAGGATGATGATGATGATGATGATGATGAT	571
Qy	541	TCATAGTCTGATAGGAGGATTAAGATTCGAGCTTTCGATGATGATGATGATGATGATGATGAT	600
Db	572	TCATATTCCTCACTACAGTATTAAGATTCGAGCTTTCGATGATGATGATGATGATGATGATGAT	631
Qy	601	AGCAAGAGTGTGGTGTGATTAAGTCTTCTCCATTTAGCAATGGGCTCTTTCACAGAACAA	660
Db	632	AGCAAGAGTGTGGTGTGATTAAGTCTTCTCCATTTAGCAATGGGCTCTTTCACAGAGAT	691
Qy	661	GGTCTCTGATAGGACCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAGGCGGAGTT	720
Db	692	GGCTCTCAGAGTGGACCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAGGCGGAGTT	751
Qy	721	GCTCACTCAATCAAGGAGGAGAGATCAAGTCTGCTGATGATGATGATGATGATGATGATGAT	780
Db	752	GCTCACTCAATCAAGGAGGAGAGATCAAGTCTGCTGATGATGATGATGATGATGATGATGAT	811
Qy	781	AACAGAGAGATTTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	840
Db	812	AATAAGATATTTCTTCCATCTGCTGCTGATGATGATGATGATGATGATGATGATGATGAT	871
Qy	841	GTTGAGAGATTTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	900
Db	872	GTTGAGAGATTTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	931
Qy	901	GAGCTATTTCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	960
Db	932	GAGCTATTTCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	991

RESULT 8	
AY264803	1221 bp mRNA linear PLN 30-APR-2003
LOCUS	Malus x domestica L-galactose dehydrogenase mRNA, complete cds.
DEFINITION	AY264803
ACCESSION	AY264803.1 GI:30267885
VERSION	
KEYWORDS	Malus x domestica (cultivated apple)
SOURCE	Malus x domestica
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.
REFERENCE	1 (bases 1 to 1221)
AUTHORS	Beuning, L., Bowen, J., Crowhurst, R., Gleave, A., Macrae, E., Newcomb, R., Perera, S., Ross, G., Snowden, K., Walton, E., and Yauk, Y.K.
TITLE	A gene database from Fruit Tree Species
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 1221)
AUTHORS	Laing, W.A. and Macrae, E.
TITLE	Kiwifruit Leaf Galactose Dehydrogenase
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 1221)
AUTHORS	Laing, W.A. and Macrae, E.

TITLE	Direct Submission
JOURNAL	Submitted (28-MAR-2003) Gene Technologies Sector, The Horticultural and Food Research Institute of New Zealand, PB 92169, Auckland, New Zealand
FEATURES	Location/Qualifiers
source	1. .1221
	/organism="Malus x domestica"
	/mol_type="mRNA"
	/cultivar="Royal Gala"
	/db_xref="taxon:3750"
	/tissue_type="pre-opened floral bud"
CDS	27. .1001
	/function="oxidizes L-galactose at C1 to L-galactono-1,4-lactone"
	/codon_start=1
	/evidence=not experimental
	/product="L-galactose dehydrogenase"
	/protein_id="AAP21783.1"
	/db_xref="GI:30267886"
	/translation="MASQPLPKVELRLNGLKSLSCVGFAGPLGNVFPVSDDEAI GSVREAPRRGNFDPSTGYCTSEKVLGKTLKALGVPRSEYVATKCGRYADGDFG SADRVTKSIDESLERLDYVDILQCHDIERGLDQIVNETIPALKKLEAGKIRFIG ITGLPIFTYVLDRIAGMVDVYLSYCHYINDSTLEDLIPYLSKSGVGLISASPLA MGLUTENGPEWHPASAEKLSACRAAAVYCKERNNISKLALQISLNDISSVLVGM NSINQVEENVAAYVELATIGKNEKILAEVEAILRPVKNQTPWPSGLQOS"
ORIGIN	
Query Match	59.4%; Score 570.2; DB 8; Length 1221;
Best Local Similarity	74.7%; Pred. No. 8.6e-153;
Matches	716; Conservative 0; Mismatches 243; Indels 0; Gaps 0;
Qy	2 TGACGAAATAGAGCTTCGAGCTTTGGGGAACACAGGGCTTAAAGGTTAGCGCGTGTGT 61
Db	43 TGCCAAAGTTGAACTCCGAGAGCTTGSCAACACAGGCTCAAGCTAGCTGTGTGGCT 102
Qy	62 TTGTGTCTTCCGCTCGAAGTGTCTTGGTTCAGTCCGCGAAGATGATGCCGTGCGCA 121
Db	103 TCGGGGCTCCCTCTCGGCAACGCTTCTGGTCCGCTCTCCGATGACGAAGCCATCGCT 162
Qy	122 CCGTCCGAGGCTTCCGCTCGGTATCAACTTCTTGACACCTCCCGCTATTATGAG 181
Db	163 CTGTTCCGGAAGCTTCCGCGAGGATCAACTTCTTGACACCTCTCCGCTATTATGAG 222
Qy	182 GAACACTGTCTGAGAAATGCTTGGTAAGGACTTAAAGGCTTTCAGTCCCTAGAGTG 241
Db	223 GGACTTTGTGAGAAAGTGTCTTGTAAAACGCTTAAAGCTTAGGTGTGCGAGAAGCG 282
Qy	242 ACTACATTTGGCTTACTAAGTGTGGTATATATAAGAGGTTTGTATTCAGTGTCTGAG 301
Db	283 AGTACGTTTGGCAACCAAGTGTGACGCTATGCTGACGGTTTGTATTCAGTGTCTGATA 342
Qy	302 GAGTAAGAAAGATTTGACGAGAGCTTGGAGAGGCTTCAGCTTGATTTATGTCATAC 361
Db	343 GAGTGACTAAGAGCATTTGATGAGAGCTTGGAGAGATTCAGCTAGATTACGTCGATATAC 402
Qy	362 TTCAATGCCATGACATTCAGTTCGGGCTCTCTTGTATCAGATTCAGTGTGAGTGAACAATTCCTG 421
Db	403 TGCAATGCCATGACATTCAGTTCGGGCTCTCTTGTATCAGATTCAGTGTGAGTGAACAATTCCTG 462
Qy	422 CTCTTCAGAACTGAAACAGAGAGGAGAGCCCGGTTCAATTCAGTGTATCATCTGCTCTTCCTG 481
Db	463 CCTTAAAGAACTGAAAGAGCTGGGAAGATTCGTTTTCATTTGGTATCATCAGGACTTCAT 522
Qy	482 TAGATATTTTTCATTTATGTTCTTGTATCAGTGTGCTCCAGGAGCTGTCCATGATATTGT 541
Db	523 TGGGAATTTTTCATATATGTTCTTGTATGATATACAGCTGGCATGTGGTGTGCTGCTGT 582
Qy	542 CATACTGTCTATTCAGGCTTAATGATTCGAGCTTGTGCTGATTTACTACCTTACTTTGAAGA 601
Db	583 CATATTGCCACTACGGTATTATTAATGATTTCAACATTTGGAGGATCTAATACCATCTCAAGA 642
Qy	602 GCAAAGTGTGGTGTGATTAAGTCTTCTCAATTAGCAATGGGCTCTCTTACAGAACAAAG 661

Db 643 GCAAGGTTGGGAATCATTAAGTCTCTCCACTTGGCAATGGGTCTGCTTACTGAGAATG 702
Qy 662 GTCTCTCTGAATGGCACCCCTGCTTCCCTAGAGCTCAAGTCTGCAAGCAAAAGCCGAGTTG 721
Db 703 GTCTCCAGAAATGGCACCCCTGATCTGCTGAAGTCTGCTTGGCCGAGCTGCTGCTG 762
Qy 722 CTCACCTGCAAAATCAAGGCAAGAGATCAAAAAGTTAGTCTGCAATACAGTTTAGCA 781
Db 763 TCTATTGTAAGAGAGAGAGGAATAATATTTCAAGTTAGTCTGCAATACAGCTTTGTCCA 822
Qy 782 ACAAGGAGATTTCCGTCGGTGTGGTGGATGAGTCTGCTCAAGGTAGAGAAATG 841
Db 823 ATAAGGATATCTCATAGTCTGGTGGCATGAATCTATTAATCAGGTTGAGGAAATG 882
Qy 842 TTGCAGCAGTTACAGAGCTTGAAGTCTGGGATGATCAAGAACTCTGCTGAGGTTG 901
Db 883 TCGCTGCTGCTGGAGCTCGCGACCATTTGGCAAGATCAGAAATTTCTAGCAGAGTTG 942
Qy 902 AAGCTATTCTCGAGCTGTAAGAAATCTGACATGAGCCAGTGGAAATCCATCAGAACTAA 960
Db 943 AAGCTATTCTGAACACAGTGAAGAAATCAGACATGCGCTAGCGGACTCCAGCAGAGCTGA 1001

RESULT 9

AB160990 1239 bp mRNA linear PLN 24-JUL-2004
LOCUS Spinacia oleracea GDH mRNA for L-galactose dehydrogenase, complete cds.
DEFINITION

ACCESSION AB160990

VERSION AB160990.1

KEYWORDS GI:50582466

SOURCE Spinacia oleracea (spinach)

ORGANISM Spinacia oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Caryophyllales; Amaranthaceae; Spinacia.

REFERENCE

1 Mieda, T., Yabuta, Y., Rapolu, M., Motoki, T., Ishikawa, T.,
Yoshimura, K. and Shigeoka, S.
Unpublished
2 (bases 1 to 1239)
Yoshimura, K. and Shigeoka, S.
Mieda, T., Yabuta, Y., Madhusudhan, R., Motoki, T., Ishikawa, T.,
Yoshimura, K. and Shigeoka, S.
Direct Submission
Submitted (23-JAN-2004) Yukinori Yabuta, Faculty of Agriculture,
Kinki University, Department of Food and Nutrition; 3327-204
Nakamachi, Nara city, Nara 631-8505, Japan
(E-mail: yabuta@nara.kindai.ac.jp, Tel: 81-742-43-7273 (ex. 3416),
Fax: 81-742-43-2252)

FEATURES

source

1..1239
/location/Qualifiers
/organism="Spinacia oleracea"
/mol_type="mRNA"
/db_xref="taxon:3562"
1..1239
/gene="GDH"
83..1051
/gene="GDH"
/function="oxidises L-galactose at C1, forming
L-galactono-1,4-lactone"
/codon_start=1
/evidence=experimental
/product="L-galactose dehydrogenase"
/protein_id="BAD32687.1"
/db_xref="GI:50582467"
/translation="WNTHQKLEERELGNLNLSCVGFASPLGNVFGDVSEEQSIAT
VIEAFNQINFPDTPYVGATLSEKVLGKLCALGASRDEYIVATKCRYAEQGFPSA
ERKVSIDSELRLEQYVDILQCHDIFGSLDQIVNETIPALQIKESGKTRFIGIT
GLPLEVYTVDPVPPGTDIVLVLSCHYCHINDLEMLPVPKSGVGVINAPLSMG
LHTENGPPEWHPASPEIKAAACKAADYCKGNKISKLALQYLSNKKDISTTLVGNMS
VKQVENVGALELETAGDKDEKTFABEINILKPIKNQSWPSGIQT"

gene

CDS

Query Match 54.0%; Score 518.8; DB 8; Length 1239;
Best Local Similarity 71.5%; Pred. No. 4.6e-137;
Matches 682; Conservative 0; Mismatches 272; Indels 0; Gaps 0;
Qy 7 AAAATAGAGCTTTCAGAGCTTTCGGGAACACAGGGCTTAAAGTTAGCGCGTGTGGTTTGGT 66
Db 98 AAATAGAGCGCCGAGAAATTCGGAAACACTGGCCCTTAACCTCAGTTGTGTGGTTTCGGC 157
Qy 67 GCCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGGAAGATGATGCGGTGCGCCACCGTG 126
Db 158 GCCTCTCTCTTGGCAAGCTTTCGGTGTATGTTTCTGAGGAACAATCCATTCGCCACTGTT 217
Qy 127 CGCAGGCTTTCGGTCTCGGTATCAACTTCTTCGACACTCCCGTATATATGAGGAACA 186
Db 218 ATTGAAGCTTTTAATCAGGGTATCAACTTCTTCGACACTTCTCCGTATCTGTTGCCACA 277
Qy 167 CTGCTGAGAAAAATGCTTGGTAAAGGACTAAAGGCTTTGCAAGTCCCTAGAAAGTCACTAC 246
Db 278 TTGTCGGAAGAAGTCTTGGGAAGTGTGTTGAAGCACTTGGTGCATCTAGAGATGAATAC 337
Qy 247 ATTGTGGCTACTAAGTGTGGTAGATATAAAGAAAGTGTGATTTTCAGTGTGAGAGATGA 306
Db 338 ATTGTGCAACGAAATGCGGCGGTTATGCCGAAGGCTTTGATTTTAGTGTGAAAGAGTG 397
Qy 307 AGAAAGATATTGACGAGAGCTTGGAGAGGCTTCAGCTTGATTTAGTTGACATCTTATCAT 366
Db 398 ACTAAGAGCATTTGATGAAAGCTTAGAAAGGTTGCAACTTGAGTATGTTGATATATCTACAA 457
Qy 367 TGCCATGACATTTAGTTCGGGTCTCTTGATCAGATTGTGAGTGAAACAATTCCTGCTCTT 426
Db 458 TGTCAATGATTTGAATTTGGTTCCTTGATCAGATTGTCAATGAGACTATTTCCGCCCTG 517
Qy 427 CAGAAATGAAACAGAGGGGAAGACCCGGTTCAATTTGGTATCACTGGTCTTCCTGGTTAGAT 486
Db 518 CAGAAATTAAGGAATCAGGGAAGACTAGATTCAATTTGGTATATAAATCGACTTCCATTGGAA 577
Qy 487 ATTTTCACTATGTTCTTGATCGAGTGCCTCCAGGAGCTGTCGATGTGATATTGTCATAC 546
Db 578 GTGTACAGCTATGTTCTTGACCGGTTTCCCGCTGGGCAAAATGATGTTGTTCTGTCATAT 637
Qy 547 TGTCAATACGGGCTTAATGATTTCGACGTTGCTGGATTTTACTACCTTACTTGAAGAGCAAA 606
Db 638 TGCCACTATTGATCAATGACTCAACCTTGGGAAGACATGCTACCGTATTTCAAGAGCAA 697
Qy 607 GGTGTGGTGTGATAAGTGTCTTCCTCATTAGCAATAGGGGCTCCTTTACAGAAACAAGTCT 666
Db 698 GCGGTTGGTGTGATCAATGCTTCTCCACTTCAATGGGGCTTCAACACAGAAAAATGGCCCC 757
Qy 667 CCTGNAATGGCACCTGCTTCCCGCTGAGCTCAAGTCTGCAAGCAAGCCGAGTTCCTCAC 726
Db 758 CCAGAGTGGCATCCAGCTTCACCTGAGATTAAAGGCTGCTGCAAAAGCTGCTGCAGATTAC 817
Qy 727 TGCAAAATCAAAAGGGCAAGAAGATCAAAAGTTAGTCTGCAATACAGTTTAGCAAAACAAG 786
Db 818 TGTAAAGAAATGGGAAGATATCTCAAAATTAGTCTGCGAGTACAGCTTGTCAAAACAA 877
Qy 787 GAGATTTTCGTGGTGTGGTGGAGTGTCTGTCTCACAGGTAGAGAAAAATGTTGCA 846
Db 878 GACATTTCCACTACTCTTTGTTGGCATGAATCTGTGTAAGCAGGTGGAGAAAAATGTTGT 937
Qy 847 GCAGTTACAGAGCTTGAAGTCTGGGGATGATCAAGAAACTCTCTCTGAGGTGGAAGCT 906
Db 938 GCGCGCTTTAGAACTTGAACCTGCAAGAAAGGATGAGAAACAATTTGCGAGAAATTTGAGAAC 997
Qy 907 ATTCTCGAGCTGTAAAGAAATCTGACATGGCCCAAGTGGAAATCCATCAGAACTAA 960
Db 998 ATCTCAAGCCCTATAAAGATCAAAAGTTGGCCCTAGTGGTATACAAACAACCTGA 1051

RESULT 10

AK060431
LOCUS Oryza sativa (japonica cultivar-group) cDNA clone:001-011-B02, full
DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:001-011-B02, full
AK060431
mRNA linear PLN 24-JUL-2003

insert sequence.
 AK060431 GI:32970449
 FLI CDNA: oligo-capping.
 ORYZA sativa (japonica cultivar-group)
 ORYZA sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.

1
 The Rice Full-length cDNA Consortium, National Institute of
 Agrobiological Sciences Rice Full-length cDNA Project Team;
 Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
 Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,
 Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,
 Ohtsuki, K., Shishiki, T., Foundation of Advancement of International
 Science Genome Sequencing & Analysis Group; Ohtsuki, Y., Murakami, K.,
 Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,
 Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,
 Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J.,
 Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,
 Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN;
 Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,
 Hara, A., Hashizume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,
 Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y.,
 Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
 Yoshino, M., and Hayashizaki, Y.

Collection, mapping, and annotation of over 28,000 cDNA clones from
 japonica rice
 Science 301 (5631), 376-379 (2003)
 2752273
 12869764

2 (bases 1 to 1234)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,
 Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T.,
 Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K.,
 Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,
 Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,
 Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,
 Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M.,
 Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,
 Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,
 Namiki, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, K.,
 Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H.,
 Osato, N., Ota, Y., Ohtsuki, R., Ohtsuki, K., Ohtsuki, K., Sakai, K.,
 Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K.,
 Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,
 Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,
 Tagawa, A., Takahashi, P., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,
 Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,
 Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and
 Yoshimura, A.
 Direct Submission
 Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of
 Agrobiological Sciences, Department of Molecular Genetics, Head of
 Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
 305-8602, Japan (E-mail:skkuchienia@affrc.go.jp,
 Tel:81-29-838-7007, Fax:81-29-838-7007)
 This clone is one of the 28K full-length cDNA clones from japonica
 rice.
 URL : <http://cdna01.dna.affrc.go.jp/cDNA/>
 NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,
 Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,
 Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,
 Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and
 Yamamoto, M.
 FAIS Genome Sequencing & Analysis Group: Ohtsuki, Y., Iida, Y.,
 Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,
 Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J.,
 Mizuno, K., Narikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S.,
 Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,
 Yoshimura, A., Matsubara, K. and Murakami, K.
 Genome Exploration Research Group in Riken Genomic Sciences Center
 and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,
 Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,
 Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,
 Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,
 Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N.,
 Ota, Y., Satoh, K., Sakai, K., Sakazume, N., Sano, H.,
 Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,
 Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,
 Yasunishi, A. and Hayashizaki, Y.

FEATURES
 Location/Qualifiers
 1..1234
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultiivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="001-011-802"

ORIGIN
 Query Match 47.3%; Score 454; DB 8; Length 1234;
 Best Local Similarity 67.1%; Pred. NO. 1.7e-118;
 Matches 643; Conservative 0; Mismatches 315; Indels 0; Gaps 0;
 QY 3 GACGAAATAGAGCTTCGAGCTTTGGGAAACACAGGGCTTAAGTTAGCGCGTGGTTT 62
 DB 65 GAGCGACATGAGCTCGCGGAGCTCGGCGCCACCGGCGCTCGCGCTCGCGCTT 124
 QY 63 TGGTGCCTCTCCGCTCGGAAGTGTCTTCGCTCCAGTCCGCGGAAGATGATCGCGCCAC 122
 DB 125 CGGCGCTCCCTCTCGGCATGCTTCGCGGAGCTCCCGCGGAGCTCGCGCGCGC 184
 QY 123 CGTGGCGAGGCTTTCCGCTTCGGTATCAACTTTTCGACACCTCCCGTATATGAGG 182
 DB 185 CGTGGCGCGCTTCGACCTCGGCTCGGCTCAACTTTTCGACACCTCCCGTATATGAGG 244
 QY 183 AACACTGCTGAGAAATGCTTGTGTAAGGAGTAAAGGCTTTGCAAGTCCCTAGAAGTGA 242
 DB 245 CACGGTTCGAGAGTCGGTTCGCGGAGCTCGCTCCGCGCGGCGGCGCGGACCG 304
 QY 243 CTACATTTGGCTACTAAGTGTGGTAGATATAAAGAGGTTTGTGATTTTCAGTGTGAGAG 302
 DB 305 GTTGTGTGCGCCACCAAGTCGGCGCTACAGGAGAGGTTTCGACTTCAGCGCGCGCG 364
 QY 303 AGTAAGAAAGATATTGACGAGAGCTTGGAGAGGCTTCAGCTTGATTTGATGACATCT 362
 DB 365 CGTCAACCGAGCTCGACGAGAGCTCGCAGCGCTCGGCGCTCGAGCTACGTCGACATCT 424
 QY 363 TCATTGCCATGACATTGATTTGGGTCTCTTGATCAGATTGAGTGAACAACTTCCTGC 422
 DB 425 CCATCGCCACACATCGAGTTTCAACCGACTCGACAGATTGTGAATGAGACGATTCGGT 484
 QY 423 TCTTCAGAACTGAAACAGAGGAGAGACCGGTTTCACTGGTATCATCTGGTCTTCCTGTT 482
 DB 485 CTTCCAGAGATCAAGAGAGCGGAGAGCGGCTTCATCGGATTAACCGGCTCCATT 544
 QY 483 AGATATTTCTACTTATGTTTCTTGATTCGAGTGCCTCCAGGAGCTGTCGATGATATGTC 542
 DB 545 GAGCATCTACACTTATGTGCTCGACAGGCTCGCGCAGGTTCTGTGGATGTGATTCGTCT 604
 QY 543 ATACTGTCATTACGGCGTTTATGATTTCGAGCTTGTGCTGATTTACTTACTTCTTGAAG 602
 DB 605 TTACTGCCATATGGGATCAATGATACCGCGCTTGTGGATTTGCTTCCCTACCTCAAGAG 664
 QY 603 CAAGAGTGTGGTGTGATAGTGTCTTCTCCATTAGCAATGGCCCTCTCTTACAGAACAGG 662
 DB 665 CAAGAGTGTGGTGTGATAGTGTCTTCTCCATTAGCAATGGGCTCTTCTACAGATACCG 724
 QY 663 TCCTCTGAATGGCACCTCTGCTTCCCTTGAGCTCAAGTCTGCAAGCAAGCGCGAGTTC 722

Db 725 GCACCCGGAATGCGACCCCTGCGACCGAAGAACTAAAGTTGGCATCGCGGCGAGCGAGA 784
Qy 723 TCACATGCAATCAAAAGGGCAAGATCACAAGTTAGCTCTGCGCAATACAGTTTACGAAA 782
Db 785 TCACATGTAAGAAGGGGAAACATTAACAAGCTAGCTATGCGATGACGCTTGATGNA 844
Qy 783 CAAGGAGATTTTCGTCGGTGTGGTGGGATGAGCTCTGTCACAGGTAGAGAAATGT 842
Db 845 CAATGAGATTTTCACGGTCTCTGTTGGAATGAATCTCCAGAACAGGTAGAGGAGATGT 904
Qy 843 TCCAGCAGTTACAGAGCTTGAAGTCTGGGATGATCAAGAACTCTCTCTGAGGTGA 902
Db 905 GGCTGCTGCAATTTGAATTTGCTACTTCGGTATTTGATAAAGAACTTCTGACGAAATGA 964
Qy 903 AGCTATTCTCGAGCCTGTAAGAAATCTGACATGCGGCAAGTCCATCCATCAGAACTAA 960
Db 965 ACAAATCTTGAGCCTGTCAGAACATGACCTGGTCCAGTGCGATGAGCAAGCCTGA 1022

RESULT 11
AK067039
LOCUS Oryza sativa (japonica cultivar-group) cDNA clone:J013091L23, full
DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:J013091L23, full
insert sequence.
AK067039
ACCESSION AK067039.1 GI:32977057
VERSION FLI CDNA; CAP trapper.
KEYWORDS Oryza sativa (japonica cultivar-group)
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

1
REFERENCE The Rice Full-length cDNA Consortium, National Institute of
AUTHORS Agrobiological Sciences Rice Full-length cDNA Project Team:
Kikuchi, S., Satoh, K., Negata, T., Kawagashira, N., Doi, K.,
Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,
Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,
Ohtsuki, K., Shishiki, T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group: Ootomo, Y., Murakami, K.,
Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,
Kurotsaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,
Narikawa, R., Sugiyama, A., Mizuno, K., Yoshimizu, S., Niikura, J.,
Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,
Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN:
Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,
Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,
Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y.,
Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Yoshino, M. and Hayashizaki, Y.
Collection, mapping, and annotation of over 28,000 cDNA clones from
japonica rice

TITLE Science 301 (5631), 376-379 (2003)
JOURNAL
MEDLINE 22752273
PUBMED 12869764
REFERENCE 2 (bases 1 to 1305)

AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,
Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T.,
Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K.,
Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,
Kawagawa, S., Kato, H., Kawagashira, N., Kawai, J., Kawamata, M.,
Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,
Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Kurotsaki, T., Kusumegi, T., Li, C., Lu, M.,
Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,
Mizuno, K., Murakami, K., Murata, M., Negata, T., Nakamura, M.,
Namiki, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, K.,
Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H.,
Osato, N., Ota, Y., Ootomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K.,
Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K.,
Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,

Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,
Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,
Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,
Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and
Yoshimura, A.
Direct Submission
Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of
Agrobiological Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki
305-8602, Japan (E-mail: skikuchi@niae.affrc.go.jp,
Tel: 81-29-838-7007, Fax: 81-29-838-7007)

COMMENT This clone is one of the 28K full-length cDNA clones from japonica
rice.

URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,
Negata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,
Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,
Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and
Yamamoto, M.

FAIS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y.,
Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,
Kodama, T., Kurotsaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J.,
Mizuno, K., Narikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S.,
Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,
Yoshimura, A., Matsubara, K. and Murakami, K.
Genome Exploration Research Group in Riken Genomic Sciences Center
and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,
Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,
Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,
Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Kanagawa, S., Kato, H., Kawai, J., Kouda, M.,
Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N.,
Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,
Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,
Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,
Yasunishi, A. and Hayashizaki, Y.

FEATURES
Location/Qualifiers

1..1305
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="J013091L23"

ORIGIN

Query Match 47.1%; Score 452.4; DB 8; Length 1305;
Best Local Similarity 67.0%; Pred. No. 4.9e-118;
Matches 642; Conservative 0; Mismatches 316; Indels 0; Gaps 0;

Qy 3 GACGAAATAGAGCTTCGAGCTTTTGGGAAACACAGGGCTTAAGTTAGCGCGTGGTTT 62
Db 50 GAGCGACATGAGCTCCGCGAGCTCGCGCCACAGGGCTCCGCGTCCGCGTGGCTT 109
Qy 63 TGGTGCCTTCGCTCGGAAGTGTCTTCGTTCCATGTCGCGGAAGATGATCGCGTCCAC 122
Db 110 CGGCGCTTCCCTCTCGGCGCATGTCTTCGGCGACGTCCCGCGACGTGCGCGCGCGC 169
Qy 123 GGTGCGGAGGCTTTCGCTCGGTATCAACTTTCGACACCTCCCGTATTTATGAGG 182
Db 170 GGTGCGGCGCGCTTCGACCTCGGATCAACTTTCGACACCTCCCGTATTTACGCGG 229
Qy 183 AACACTGTCTGAGAAATGCTTGGTAAAGGACTTAAAGGCTTTGCAAGTCCCTAGAGTGA 242
Db 230 CAGCGTGTGGAGTGGTGTCTCGGGAGTCCGCTCCGCGCGCGGCGTCCGCGGACCG 289
Qy 243 CTACATTTGGTGTCTTAAGTGTGGTAGATATAAAGAGGTTTGTATTTCAGTCTGAGAG 302
Db 290 GTTCGTGTGCGCCACCAAGTGGCGGCTTACAGGGAAGGGTTTCGACTTCAGCGCGCGC 349
Qy 303 AGTAAGAAGAGTATTGACGAGGCTTCGAGAGGCTTCAGCTTGTATTATTGTCATACT 362

350	CGTCA	CCCCAGCGTCGACGAGAGCCTCGACGCGCTCGAGCTCGACTACGCTCGACATCTCT	409
363	TCATTGCCATGACATTTGAGTTGGGGTCTCTTTGATCAGATTGTGAGTGAAACAAATTCCTCGC	422	
410	CCACTGCCACGACATCGAGTTTCAACGACCTCGACCGAGTTGTGAATGAGACGATTCGGGT	469	
423	TCTTCAGAACTGGAACAGAGAGGGGAGACCGGGTTCATTTGGTATCACTGGTCTTCGGTT	482	
470	GCTCCAGAAGATCAAGGAGAGCGGAGAGCGCGGTTTATCGGGATAACCGGGCTGCGATT	529	
483	AGATATTTTTCACTTATGTTCTTTGATCGAGTGCCTCCAGGGACGTGTCGATGTGATATGTGC	542	
530	GAGCATCTACATTTATGTGCTCGACACAGTGCGCCAGGTTCTGTGGATGTGATTCGTGC	589	
543	ATACTGTCAATTACGGGGTTAATGATTCGACGTTGCTCGATTTACTACCTTACTTTGAAGAG	602	
590	TTACTGCCACTATGGGATCAATGATACCGCGCTTTGTGGATTTGCTTCCCTACCTGAAGAG	649	
603	CAAAGTGTGGGTGTGATAAGTGCTTTCTCCATTAGCAATGGCGCTCCCTTACAGAAACAAGG	662	
650	CAAAGTGTGGCGTGATCAGTGCTTCACCTCTTGGCAATGGGGCTTCTTACAGATAACGG	709	
663	TCCTCTGAAATGGCACCTCTGCTTCCCTTGAGCTCAAGTCTGCAAGCAAAAGCCGAGTTGC	722	
710	GCCACCGAATGGCACCCCTGCACCGAAAGAACTAAAGTTGGCATGCGGGCAGCAGCAGA	769	
723	TCACTGCAAAATCAAAAGGGCAAGAAGATCAAAAGTTAGCTCTGCAATACAGTTTTCAGAA	782	
770	TCACTGTAGAGAAAGGGAAACAACTTACAAAGCTAGCTATGCAGTACAGCTTGTATGAA	829	
783	CAAGGAGATTTCTCGTGGTGTGGTTGGGATGAGCTCTGTCTCAAGGTAGAGAAATGT	842	
830	CAATGAGATTTTCCACGGTCTTTGTGGAAATGAACTCTCCAGAAACAGGTAGAGGAGAAATGT	889	
843	TGCAGCAGTTACAGAGCTTGAAAGTCTCGGGATCGATCAAGAACTCTGTCTGAGGTTCG	902	
890	GGCTGCTGCAATTGAAATTTGTCTACTTCGCGTATTGTATAAGAACTTCTGCACGAGTTGA	949	
903	AGCTATTTCTCGACGCTGTAAAGAAATCTGCATGCGCCAAAGTGGAAATCCATCAGAACTAA	960	
950	AGCAATTTCTGAGCTGTGCAAGAACATGACCTGGTCCAGTGGCAATTCGAGAGCGCTGA	1007	

Saito,R., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Yoshino,M. and Hayashizaki,Y.
Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice
Science 301 (5631), 376-379 (2003)
22752273
12863764
2 (bases 1 to 1275)
Adachi,Y., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K., Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K., Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I., Kanegawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M., Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M., Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Kurotaki,T., Kusumegi,T., Li,C., Lu,M., Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A., Mizuno,K., Murakami,K., Murata,M., Negata,T., Nakamura,M., Namiki,T., Narikawa,R., Nikura,Y., Nishi,K., Nomura,K., Onodera,S., Ohtsuka,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H., Oeato,N., Ota,Y., Otsu,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K., Shingawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Taya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and Yoshimura,A.
Direct Submission
Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)
This clone is one of the 28K full-length cDNA clones from japonica rice.
URL : <http://cdna01.dna.affrc.go.jp/cdna/>
NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Onodera,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and Yamamoto,M.
FAIS Genome Sequencing & Analysis Group: Otsu,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurotaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Nikura,Y., Oka,M., Ryu,R., Sugano,S., Sugiyama,A., Satsubara,K. and Murakami,K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanegawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Oeato,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Taya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.
Location/Qualifiers
1. 1275
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="J033087M07"

Best Local Similarity 66.9%; Pred. No. 1.4e-117; Matches 641; Conservative 0; Mismatches 317; Indels 0; Gaps 0;			
QY	3	GACGAAATAGAGCTTCGAGCTTCGGGAAACACAGGCTTAAGGTTAGCGCGTGTGTTT	62
Db	27	GAGCAGATGAGACTCCGGAGCTCGCGCCACAGGCGCTCCGCTCAGCCCGTCGGTT	86
QY	63	TGTCCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGCAAGATGATCGCGTCCGAC	122
Db	87	CGCGCCCTCCCTCTCGGCCATGTCTTCGGGACGCTGCCCGCAGCTCGCCCGCGCGC	146
QY	123	CGTCGGAGGCTTCGCTCGGTATCAACTCTTCGACACTCCCGTATTATGAGG	182
Db	147	CGTCGGCGCGCCCTCGACTCGGCATCAACTTCTTCGACATCCCGTACTACGGCG	206
QY	183	ACACTGTCTGAGAAATCTTGTGTAAGGAGCTTAAAGGCTTTGCAAGTCCCTAGAAGTGA	242
Db	207	CAGGTGTGCGAGTGGTGCTCGGGACTGCTCCGCGCGCGGCTGCGCGGACCG	266
QY	243	CTACATTGTGGCTACTAAGTGTGGTAGATATAAGAGGTTTGTATTTCACTGCTGAG	302
Db	267	GTTCTGTGTGCGCACCAAGTGGCGCTACAGGAAGGTTTCGACTTCAGCGCGCGCG	326
QY	303	AGTAAGAAGATATTGACGAGCTTGGAGAGGCTTCAGCTTGATTTGATGATCTACT	362
Db	327	CGTCACCGCAGCGTCGACGAGGCTTCGACGCTTCGCGCTTCGACTCGACATCCT	386
QY	363	TCATTGCCATGATTTGAGTTCGGGTCTCTTCATCAGATTGTGAGTGAACAATTCCTGC	422
Db	387	CCACTGCCACGATCGAGTTACCGACTCGACGATTTGATGATGACGATTCGGT	446
QY	423	TCTTCAGAACTGAACAAGAGGGAAGACCGGTTTCATTTGATTCATCTGCTTCCTG	482
Db	447	GCTCAGAGATCAAGGAGCGGGAAGCGCGGTTTCATCGGGAATAACCGGCTGCCAT	506
QY	483	AGATATTTTCACTTATGTTCTGATCGAGTGCCTCAGGACTCTCGATGATATTGTC	542
Db	507	GAGCATCTACATTTATGTCTCGACGAGTGCCTCGAGGTTCTGTGGATGATTTCTGTC	566
QY	543	ATACTGTCAATTCGGGTTTAATGATTGCGAGTTCGCTGCTGATTTACTTACTTCAAGAG	602
Db	567	TTACTGCCACTATGGATCAATGATACCGGCTTGTGGATTTGCTTCCCTACTGAGAG	626
QY	603	CAAGGTGTGGTGTGATAGTCTTCTCCATAGCAATGGCCCTCCTTACAGAACAGG	662
Db	627	CAAGGTGTGGGTGATCAGTGTTCCTCTCTGCAATGGGCTTCTTACAGATAACGG	686
QY	663	TCTCTGAAATGCGACCTCTGCTCCCTGAGCTCAAGTCTGCAAGCAAGCCGAGTTGC	722
Db	687	GCCACCCGAATGCACTCCCTGCAAGCAAGCAAGTTCGAGTTCGAGGCGAGCAGA	746
QY	723	TCATGCTCAATCAAGGGCAAGAGATCAAAAGTTAGTCTGCAATACAGTTTAGCAA	782
Db	747	TACTGTGAAGAGAGGGGAAAAACATTAAGCTAGTATGCGATGACGTTGATGNA	806
QY	783	CAAGGAGATTTCTCGGTGTTGTTGGATGAGCTCTGCTCAGAGTAGAAGAAATGT	842
Db	807	CAATGAGATTTCCACGGTCTTGTGTTGGAATGAATCTCCAGAACAGGTAGAGGAAAT	866
QY	843	TCGACGAGTTACAGAGCTTGAAGTCTGGGATGATCAAGAACTCTCTCTGAGTTGA	902
Db	867	GGCTGCTGAATTTGAATTTGCTTCTGCTGCTGATTTGATAAAGAACTCTTCGACGA	926
QY	903	AGCTATTTCTCGAGCTGTAAAGATCTGACATGGCGCAAGTGAATCCATCAGAACTAA	960
Db	927	AGCAATTTCTGAGCTGTCAAGACATGACCTGGTCCAGTGGCATTTGACCAAGCCTGA	984

RESULT 13
ATT16L1/c
LOCUS
DEFINITION Arabidopsis thaliana DNA 98124 bp DNA linear PLN 27-AUG-1999
project).

ACCESSION VERSION KEYWORDS SOURCE ORGANISM	AL0311394 AL0311394.1 GI:3859658 Arabidopsis thaliana (chale cress) Arabidopsis thaliana Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
REFERENCE AUTHORS	1 Bevan, M., Obermaier, B., Deutschenbaur, S., Piravandi, E., Hoheisel, J., Jesse, T., Heijnen, L., Vos, P., Mewes, H. W., Mayer, K. F. X. and Schueller, C. unpublished EU Arabidopsis sequencing, project. Direct Submission Submitted (09-NOV-1998) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UU Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk On Nov 11, 1998 this sequence version replaced gi:3549653.
COMMENT FEATURES	Location/Qualifiers 1..98124 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4" 1..6620 /gene="T16L1.10" Join(<1..228,315..368,536..689,1974..2048,2138..2415, 2754..2981,3156..3254,3388..3537,3662..3808,4065..4115, 4197..4354,4506..4638,4724..4821,5206..5452,5701..5904, 6212..6292,6393..6620) /gene="T16L1.10" note="Contains E1-E2 ATPases phosphorylation site [DKTGLT], Heavy-metal-associated domain [VGMTCGCSASVKILSQPVASAVNL]" /codon_start=1 /product="metal-transporting P-type ATPase (fragment)" /protein_id="CAA20565.1" /db_xref="GI:3549654" /db_xref="GOA:Q9SZC9" /db_xref="UniProt/Swiss-Prot:Q9SZC9" /translation="AVLPVIRHLECLSSSPSPRSISSGGSGFGYNGSGGGGG GSESDSKSLGANASDGVSPSSDIILDVGMTCGCSASVKILSQPVASAV NLTTETALVWPVPEAKSVPMOKSIGETLANHLCNCFQSTPRDLVTNFPKVPFTKT KDQARLKESRELAWSALCAVCLVGHLEHFLGVNAPMIIHISTGFHVSICLITLL GPKRLVLDDGIKSLKSPNNMTLVGLGALSFSVSSLAAMIPLKWKTFPEEPWMLI AFVLGRNLEQRAKIKATSDMTGLLSVLPKARLLDGLQNSTVEVPCNSLSVGDLY VILPGRVPADGVKSGRSTIDESFTGEPLFTVKBSGSQVAAGSINLNGLTVEVHR SGGTAVGDIIRLVEEAQSRAPVQQLVDKVGAFYGVGMALSAATFTFWNLFGAHL PSALHNGSPMSIALQLSCSLVAVCPALGLATPTAMLVGTSLGARRGLLRGGDILE KPSLVDTVPDKTGLTKGHPVTEVIIIPENPRNLNDTWESEVLMIAAAVESNTH PVGAIVKAARARNCTMKADSDTFEFGSGAVAIYNNKRVTVGTLEWKRKHGTGN SLLLALBEHINNVSIVYIGVDNTLAAVIRFEDKVEDAAQVVENLROGIDIVMLSGD KRNAANYAVSVGNHVERIVAGVPAEKKNFINELQKNKIKIVAMYGDDGINDAALAS NVGAMGGAGAGSESPVLMGNRLTQLDAMELSRQTMKTMTKQNLWFAFGYNTVGI PIAAGVLPLGTMTLTPSMAGALMGVSSLSGVTNLSLLRLYRFFSNRNDQNVKPEKEG TKQPHENTRWKSS"
exon	1..228 /gene="T16L1.10" /number=1 229..314 /gene="T16L1.10" /number=1 315..368 /gene="T16L1.10" /number=2 369..535 /gene="T16L1.10" /number=2
intron	
exon	
intron	

exon 536. .689
/gene="T16L1.10"
/number=3
intron 690. .1973
/gene="T16L1.10"
/number=3
exon 1974. .2048
/gene="T16L1.10"
/number=4
intron 2049. .2137
/gene="T16L1.10"
/number=4
exon 2138. .2415
/gene="T16L1.10"
/number=5
intron 2416. .2753
/gene="T16L1.10"
/number=5
exon 2754. .2981
/gene="T16L1.10"
/number=6
intron 2982. .3155
/gene="T16L1.10"
/number=6
exon 3156. .3254
/gene="T16L1.10"
/number=7
intron 3255. .3387
/gene="T16L1.10"
/number=7
exon 3388. .3537
/gene="T16L1.10"
/number=8
intron 3538. .3661
/gene="T16L1.10"
/number=8
exon 3662. .3808
/gene="T16L1.10"
/number=9
intron 3809. .4064
/gene="T16L1.10"
/number=9
exon 4065. .4115
/gene="T16L1.10"
/number=10
intron 4116. .4196
/gene="T16L1.10"
/number=10
exon 4197. .4354
/gene="T16L1.10"
/number=11
intron 4355. .4505
/gene="T16L1.10"
/number=11
exon 4506. .4638
/gene="T16L1.10"
/number=12
intron 4639. .4723
/gene="T16L1.10"
/number=12
exon 4724. .4821
/gene="T16L1.10"
/number=13
intron 4822. .5205
/gene="T16L1.10"
/number=13
exon 5206. .5452
/gene="T16L1.10"
/number=14
intron 5453. .5700
/gene="T16L1.10"
/number=14
exon 5701. .5904

intron /gene="T16L1.10"
/number=15
exon 5905. .6211
/gene="T16L1.10"
/number=15
intron 6212. .6292
/gene="T16L1.10"
/number=16
exon 6293. .6392
/gene="T16L1.10"
/number=16
gene 6393. .6620
/gene="T16L1.10"
/number=17
gene 7274. .11125
/gene="T16L1.20"
complement(join(7274. .7811,7900. .8422,8677. .8931,
9058. .9166,9265. .9317,9422. .9682,10151. .10429,
10510. .10738,10832. .11125))
/gene="T16L1.20"
CDS complement(join(7274. .7811,7900. .8422,8677. .8931,
9058. .9166,9265. .9317,9422. .9682,10151. .10429,
10510. .10738,10832. .11125))
/gene="T16L1.20"
/note="contains EST gb:AA585990, AA597997, T13770, W43869"
/codon_start=1
/product="putative potassium transporter AtKT5p (ACKTS)"
/protein_id="CAA20566.1"
/db_xref="GI:3859659"
/db_xref="GOA:Q9SU96"
/translation="MFHVESSESGDSEIDEEFGDDSTTSLRWVDEKDDYEVNE
DYDDGDYDEHNHPMDSDDEDNDVEQRLIRTSFPAVDSDFDALRIPTQKNEIEDTGI
GKLIILALOTLGVVFGDIGTSPVFTVMPRRSPINDKEDIIGALSIVITLILPLV
KYHFVLWANDDEGGTFFALYSLICRHANVSLIPQLPSDARIISGRLKVPSPERS
LIIRERLEASWALKKLLILVIAGTAMVIADAVTIPMSGTMLSGYIFQVQVVISVS
PLVILFSVQYKTSKGLVLGVPALLMFFLAGIGIYNLVKIDYSSVFKAFNPAYIYFF
FKRNSNVAVALGCVLCATGSEAMFADLSYFVSHISQLTFLILVPLCLLGLGQAA
YLSENFQRCRCFLVSLFWPFLISNVAALIASRAMTTATFTCIQSIALGCFPRL
KIHTSKFIQIYIVPLNWSLLVCLIVGCTSNIPAINAGYIAEIGIMWTTILV
TLIMLLIQWNIIVSMFATVSLIVELVFSSVCSVADGSMIILVFATIMFIMFVN
NYGKGIYETEVQKLLPMDLLRELGSNLGTRAFGLLNYNELAKGVPAIFGHFTTL
PAIHSWIFVICIKYQVPVSPQTERFLFRVCPRSYHLFRCAVRYKDVKRESHOAF
EQILIESLEKPIRKEAQRALSDGNDTSEDITLRLVLIAPNGSVSILGVPLLA
EHMNSNKRPMERKASIDFGAGPSALDVEQSLKELSFTHKAKESGVVYLLGHGDI
RATKDSWFLKXKLVINLYAFLRKNSRRGITNLVSPHTLMQVGMTYMY"
complement(7274. .7811)
/gene="T16L1.20"
/number=1
intron complement(7812. .7899)
/gene="T16L1.20"
Query Match 45.3%; Score 435; DB 8; Length 98124;
Best Local Similarity 86.3%; Pred. No. 8.9e-113;
Matches 529; Conservative 0; Mismatches 0; Indels 84; Gaps 1;
Qy 171 GTATTATGGAGAACACTGTCTGAGAAATCTTGGTAAGGACTTAAAGGCTTTTCCAAGT 230
|||||
Db 51886 GTATTATGGAGAACACTGTCTGAGAAATCTTGGTAAGGACTTAAAGGCTTTTCCAAGT 51827
|||||
Qy 231 CCCTAGAAGTGACTACATTGTGGCTACTAAGTGTGGTAGATATAAAGAAGGTTTGTGATTT 290
|||||
Db 51826 CCCTAGAAGTGACTACATTGTGGCTACTAAGTGTGGTAGATATAAAGAAGGTTTGTGATTT 51767
|||||
Qy 291 CAGTGTGAGAGAGTAAAGAGATATTGACGAGAGCTTGGAGAGGCTTCAGCTTGATTA 350
|||||
Db 51766 CAGTGTGAGAGAGTAAAGAGATATTGACGAGAGCTTGGAGAGGCTTCAGCTTGATTA 51707
|||||
Qy 351 TGTTGACATACCTTCATTGCCATGACATTGAGTTCGGGTCTCTTGAT ----- 396
Db 51706 TGTTGACATACCTTCATTGCCATGACATTGAGTTCGGGTCTCTTGATCAGGTTTACCACCTT 51647
|||||
Qy 397 ----- 396

	gene	5401..7841 /gene="AT4g33680" complement(join(5401..5694,5785..5991,6067..6387, 6489..6539,6625..6729,6942..7004,7289..7405,7601.. 7787..7841)) /gene="AT4g33680" complement(join(5401..5694,5785..5991,6067..6387, 6489..6539,6625..6729,6942..7004,7289..7405,7601.. 7787..7841)) /genes="AT4g33680" /notes="similarity to aminotransferase (AspC family), Aquifex aeolicus, PIR2:D70479 Contains Aminotransferases class-I pyridoxal-phosphate attachment site AA267-280;Prokaryotic membrane lipoprotein lipid attachment site AA312-322"
CDS		/codon_start=1 /product="putative protein" /protein_id="CAH80085.1" /db_xref="GI:7270317" /db_xref="GOA:O81885" /db_xref="UniProt/TREMBL/O81885" /translation="MAKRVNTCKVAIPQEKLEYTKVSRNSNMSKLQAGYLPPIAR RRAAHKTYPDAQLVISLGIDTTEPIPEVITSAAMAKHAELSTISGYSYGAEQGAKP LRAAIAKTFYGLGGDDDFVSFGDKACDISRLQMFGSNVTIAVDPSYPAYDVSSV IMGQTGFNTDVKYINBYMCTPENGFPPDLSTVGRTDIIFCGSPNNPTGAATRE QLTOLVEFAKKNGSIIVYDSAAYMTSDNDPRSIFEIPGAEBVAMETASFVKYAGFTG VRLGTWVIPKLLLYSDGFPVAKDNRIICTCFNGASNISQACALACLTPGLEAMHKV IGFYKENTNIILDTFSLGYDYGVKGKNAPYVMHFEPNQSSWDVPFAEILEKTHVTTTPG SGFGPGGEGFVRVSPAGPHRENILEACRRPKQLYK"
exon		complement(5401..5694) /genes="AT4g33680" /number=1 complement(5695..5784) /gene="AT4g33680"
intron		/number=1 complement(5785..5991) /genes="AT4g33680"
exon		/number=2 complement(5992..6066) /genes="AT4g33680"
intron		/number=2 complement(6067..6387) /genes="AT4g33680"
exon		/number=3 complement(6388..6488) /genes="AT4g33680"
intron		/number=3 complement(6489..6539) /genes="AT4g33680"
exon		/number=4 complement(6540..6624) /genes="AT4g33680"
intron		/number=4 complement(6625..6729) /genes="AT4g33680"
exon		/number=5 complement(6730..6941) /genes="AT4g33680"
intron		/number=5 complement(6942..7004) /gene="AT4g33680"
exon		/number=6 complement(7005..7288) /genes="AT4g33680"
intron		/number=6 complement(7289..7405) /genes="AT4g33680"
exon		/number=7 complement(7406..7600) /gene="AT4g33680"
intron		/number=7 complement(7601..7668) /gene="AT4g33680"

```
/number=8
complement (7669. .7786)
/gene="AT4g33680"
/number=8
complement (7787. .7841)
/gene="AT4g33680"
/number=9
8752. .9691
/gene="AT4g33690"
Join (8752. .8920, 9015. .9691)
/gene="AT4g33690"
/codon start=1
/product="hypothetical protein"
/protein_id="CAB80086.1"
/db_xref="GI:7270318"
/db_xref="UniProt/TREMBL:O81886"
/translation="MDLTENRIASILLREAAELRRQAEKDGVRAYLEKPNVRRPNS
RFLATVLGVQOANKAVETNEMSLRSKEIEFGRLEKKSDESSNGOSECNNRDEL
KRCTSDENVYTTLSLSPSSRSRKRWQSEDDDEGLDGVKFLQSRYPHYTQMLP
DYFLYNDRIRIDRLLLIIVLVYDRVGRGSGVAMMDEPLPCLPERLSRTSDTG
RKLVIQPSRPLRRRTDSSDEEVYKRAHRRKKEHKKLKKHKSKEKRRDKKR
KYGRD"
8752. .8920
/gene="AT4g33690"
/number=1
8921. .9014
/gene="AT4g33690"
/number=1
9015. .9691
/gene="AT4g33690"
/number=2
10101. .12658
/gene="AT4g33700"
complement (Join (10101. .10145, 10220. .10491, 10571. .10686,
10760. .10833, 10960. .11047, 11157. .11296, 11383. .11469,
11562. .11675, 11811. .11939, 11982. .12058, 12565. .12658))
/gene="AT4g33700"
complement (Join (10101. .10145, 10220. .10491, 10571. .10686,
10760. .10833, 10960. .11047, 11157. .11296, 11383. .11469,
11562. .11675, 11811. .11939, 11982. .12058, 12565. .12658))
/gene="AT4g33700"
Note="similarity to hemolysin, 38K, Synecocystis sp.,
Pir2:S7595
contains EST gb:AI999192.1"
/codon start=1
/product="putative protein"
/protein_id="CAB80087.1"
/db_xref="GI:7270319"
/db_xref="UniProt/TREMBL:O81887"
/translation="MSGILTLGMSLSLVDELVLAKSGTPPEHRKYAAKILPVVKQHL
LVTLLIENAAAEVSGMVFDSDLSFQTLFLDGLVTAWGAILISVTLILFGEII
QSFICSRGLAIGATVAPFVPLVFCIPVAPISKLADFLGHRRAALPRABKTL
VDHGNAGKGBELTHDETTIAGALELSEKWKDAMTPISDIFIDINAKLDRLAN
LILEKHSRVPVYEQPNIIIGLVLYKNLLINPDBEIPVKNVIRIRIPRVEILPLY
DILNEFQGLSHMAVVQCDKIHPKSPKSGVKEARVDVDESGTTPQERMLRTKRS
LQKWSFPNRASSFKGSKSKWKNDADILQNLGNPLPKLBEAEAVGITMEDVI
BELLQEBIFDETTHHFEDS"
complement (10101. .10145)
/gene="AT4g33700"
/number=1
complement (10146. .10219)
/gene="AT4g33700"
/number=1
complement (10220. .10491)
/gene="AT4g33700"
/number=2
complement (10492. .10570)
/gene="AT4g33700"
/number=2
complement (10571. .10686)
/gene="AT4g33700"
/number=3
complement (10687. .10759)
```

```
/number=3
complement (10760. .10833)
/gene="AT4g33700"
/number=4
complement (10834. .10959)
/gene="AT4g33700"
/number=4
complement (10960. .11047)
/gene="AT4g33700"
/number=5

Query Match 45.3%; Score 435; DB 8; Length 192861;
Best Local Similarity 86.3%; Pred. No. 9.8e-113;
Matches 529; Conservative 0; Mismatches 84; Gaps 1;

QY 171 GTATTATGGAGAACACTGTCTGAGAAATCCTTGGTAAGGACCTAAAGGCTTTGCAAGT 230
Db 4668 GTATTATGGAGAACACTGTCTGAGAAATCCTTGGTAAGGACCTAAAGGCTTTGCAAGT 4609
QY 231 CCCTAGAAGTGACTACATTGTGGCTACTTAAGTGTGGTAGATATAAAGAGGTTTGTATTT 290
Db 4608 CCCTAGAAGTGACTACATTGTGGCTACTTAAGTGTGGTAGATATAAAGAGGTTTGTATTT 4549
QY 291 CAGTCTGAGAGAGTAGAAGAGATATTGACGAGAGCTTTGAGAGGCTTCAGCTTGATTA 350
Db 4548 CAGTCTGAGAGAGTAGAAGAGATATTGACGAGAGCTTTGAGAGGCTTCAGCTTGATTA 4489
QY 351 TGTGTGACATCTTCATTGCCATGACATTGAGTTCGGGTCTCTTGAT ----- 396
Db 4488 TGTGTGACATCTTCATTGCCATGACATTGAGTTCGGGTCTCTTGATCAGGTTTACCACCTT 4429
QY 397 ----- 396
Db 4428 TATTTCCTTGGGTTTATCACTTTTCGTTTGTGTCTCAAAACCTTTTACTAAGTGTTCAT 4369
QY 397 -----CAGATTGTGAGTGAACAATTCCTGCTCTTCAGAACTTGAACAAGAGGG 446
Db 4368 TCATTGTGCTGCCAGATTGTGAGTGAACAATTCCTGCTCTTCAGAACTTGAACAAGAGGG 4309
QY 447 GAAGACCCGGTTCATTGTGTATCACTGGTCTTCGTTAGATATTTTCACTTATGTTCTTGA 506
Db 4308 GAAGACCCGGTTCATTGTGTATCACTGGTCTTCGTTAGATATTTTCACTTATGTTCTTGA 4249
QY 507 TCGAGTGTCTCCAGGAGCTGTGATGTGATATTGTCTACTGTCTATTAACGGCTTAATGA 566
Db 4248 TCGAGTGTCTCCAGGAGCTGTGATGTGATATTGTCTACTGTCTATTAACGGCTTAATGA 4189
QY 567 TTCGAGTGTCTGGATTACTACCTTACTTGAAGAGCAAGGTGTGGGTGTGATTAAGTGC 626
Db 4188 TTCGAGTGTCTGGATTACTACCTTACTTGAAGAGCAAGGTGTGGGTGTGATTAAGTGC 4129
QY 627 TTCTCCATTACCAATGGSCCTCCTTACAGAACAAAGGTCTCTCTGAATCGCACCTGCTTC 686
Db 4128 TTCTCCATTACCAATGGSCCTCCTTACAGAACAAAGGTCTCTCTGAATCGCACCTGCTTC 4069
QY 687 CCCTGAGCTCAAG 699
Db 4068 CCCTGAGCTCAAG 4056
```

RESULT 15

```
ATCHRIV79/c
LOCUS ATCHRIV79 199536 bp DNA linear PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 79.
ACCESSION AL161583
VERSION AL161583.2 GI:7270260
KEYWORDS
SOURCE
ORGANISM Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
```

REFERENCE AUTHORS	1 (bases 1 to 65668) Peters,S.A., van Staveren,M., Dirkse,W., Stiekema,W., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.	exon	AAGTTDLADGCGAYLTRFCQNDKIVVAVDIDEDMELLSQRRSRQNAIKETHTLEWIEI HFPGFLFQIHPGHNHFALHGSRPKSEICRSFGAEIILDDNPRIABECANIGMKVLLFD YENSYPWSKTESVDHRPLVTRVHNWEEVEQIILSLAVSKC"
JOURNAL REFERENCE AUTHORS	2 (bases 64780 to 161254) Rose,M., Hempel,S., Entian,K.-D., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.	intron	/gene="AT4g33140" /number=1 10208. .11381 /gene="AT4g33140" /number=1 11382. .11406 /gene="AT4g33140" /number=2 11407. .11764 /gene="AT4g33140" /number=2 11765. .11899 /gene="AT4g33140" /number=3 11900. .12012 /gene="AT4g33140" /number=3 12013. .12223 /gene="AT4g33140" /number=4 13567. .19157 /gene="AT4g33150" complement(join(13567. .13619,13721. .13790,13885. .14046, 14147. .14255,14352. .14583,14665. .14823,14920. .14999, 15101. .15230,15354. .15471,15624. .15689,15780. .15860, 15955. .16305,16509. .16649,16840. .16900,16989. .17155, 17246. .17342,17426. .17597,17730. .17805,17899. .18015, 18084. .18166,18247. .18410,18494. .18633,18712. .18798, 18879. .19157)) /gene="AT4g33150" complement(join(13567. .13619,13721. .13790,13885. .14046, 14147. .14255,14352. .14583,14665. .14823,14920. .14999, 15101. .15230,15354. .15471,15624. .15689,15780. .15860, 15955. .16305,16509. .16649,16840. .16900,16989. .17155, 17246. .17342,17426. .17597,17730. .17805,17899. .18015, 18084. .18166,18247. .18410,18494. .18633,18712. .18798, 18879. .19157)) /gene="AT4g33150" note="contains EST gb:W43474, T13618, AA585912, N37568, T04246, T45802, T88051"
JOURNAL REFERENCE AUTHORS	3 (bases 147320 to 199536) Obermaier,B., Deutschenbaur,S., Piravandi,E., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.	exon	/product="lysine-ketoglutarate reductase/saccharopine" /protein_id="CAB80032.1" /db_xref="GI:7270263" /db_xref="GOA:Q9SNZ4" /db_xref="UniProt/TREMBL:Q9SNZ4" /translation="MNSNGHEEEKLGNVVGIIAETVKNWERTTLPSPHCARLLHG KDRGTGIRIVQPSAKRIHHDALYEDVGCISDDSLDCGLILGKQPELEMILPERA YAFSGTHKAQKNPDLIDKILSERVTLCDYELIVDGHGKRLLAFGKTAGRAGLVDFL HGIGORKLILGYSTPFLSLGASVMYSLAAKAAVSVGERIASOGLPGICPLVPLVF TGTGNVSLGAQEI EKLLPHTTVEPSKLPFLVKDGI SQNGISIKRVVQVTCIITSQ DWERKDPKSFDRKADYIAHEPHNPVFHEKISPYTSVLVNCMYWEKFPCLLSYKQL QDLTKGLPLVIGICDITIGGSIEFVNRATLIDSPFPFRPNNSYFDDMDGVLG MAVDILPTPEAKASQHFGLDSGVSLASMTSEISDLPALKRACISYRGELTSLYE YIPMRKSNPEEAQNDIANGVSQRTFNILVSLSHGLFDKFLINEADMDIABAQSGF HLAKCELQSDAESYSLEVGADDKRVLDQI IDSLTRLANPNEDIYISPHREANKISL KIGKVQGENEIKKPEMTKKSGLVILGAGRVCRAADFLASVRTTSSQOYKTYFGAD SEKTDVHVIVASLYLKDAKTVEGISVEAVRLDVSDESILLKYVSQDVVLSLLPA SCHAVAKTCIELKKHLVTASVDDETSMLEHKAQKASGITILGEMGLDPIGIDHMMAMK MINDAHIKKGVKSFTSYCGGLPSPAAANNPLAYFSWNPAGATIGENGLDPIGIDHMMAMK IIHVQGNLYDASAARFNPALFALCEFPNDRSLVYGHYIESEATTIFRGTLRYE GFSMATLSKLGFFDSANOVLTGKRITFGALLISNLINKDANSEATPLAGEEIEISK RIIKLGHSKETAAKAAKTIVELGFNEEREVPSLCKSVFATCYLMEELKAYSGNEQDM VLLHHEVEVEPLESKRIEKKHTATLLEFGDIKNGQTTTAMAKTVGIPAAIGALLLIEDK IKTRGVLRPLEAVYLPALDILQAYGIKMEKAE" complement(13567. .13619) /gene="AT4g33150" /number=1 complement(13620. .13720)
JOURNAL REFERENCE AUTHORS	4 (bases 1 to 199536) EU Arabidopsis sequencing,project. Direct Submission Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemckemips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk	intron	
COMMENT	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/ this fragment has an overlap with ATCHRIV78 at the 5' end and an overlap with ATCHRIV80 at the 3' end.	gene	
FEATURES source	1. 199536 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4" 7525. .8075 /gene="AT4g33130" complement(join(7525. .7590,7667. .7852,7950. .8075)) /gene="AT4g33130" complement(join(7525. .7590,7667. .7852,7950. .8075)) /gene="AT4g33130" /codon_start=1 /product="hypothetical protein" /protein_id="CAB80030.1" /db_xref="GI:7270261" /db_xref="UniProt/TREMBL:Q9SNZ6" /translation="MIQIKRLOOQLEBINLRALTSVAHESSSPFMDSPCELPDKEL LDLSAILETVSKLQESVSLRYLRQEKNERRLSEILQKSHYSAPSKFTNAQFPFN KSVTRKRSKQVDSLDSALQVE" complement(7525. .7590) /gene="AT4g33130" /number=1 complement(7591. .7666) /gene="AT4g33130" /number=1 complement(7667. .7852) /gene="AT4g33130" /number=2 complement(7853. .7949) /gene="AT4g33130" /number=2 complement(7950. .8075) /gene="AT4g33130" /number=3 9802. .12223 /gene="AT4g33140" join(9802. .10207,11382. .11406,11765. .11899,12013. .12223) /gene="AT4g33140" /codon_start=1 /product="hypothetical protein" /protein_id="CAB80031.1" /db_xref="GI:7270262" /db_xref="UniProt/TREMBL:Q9SNZ5" /translation="MSSFTDPLFLRLSYSPRSIAIRFHGDFAHRLYNPIQIYNNNNN SGSNGLPKVGCGKPGKGTFLRSCTFDSHKKPDQKARALTQHRRFLHDLGSGSERA		
gene		CDS	
gene		exon	
intron		intron	
exon		exon	
intron		intron	
exon		exon	
gene		exon	
CDS		intron	

/gene="AT4g33150"
/number=1
exon complement (13721. .13790)
/gene="AT4g33150"
/number=2
intron complement (13791. .13884)
/gene="AT4g33150"
/number=2
exon complement (13885. .14046)
/gene="AT4g33150"
/number=3
intron complement (14047. .14146)
/gene="AT4g33150"
/number=3
exon complement (14147. .14255)
/gene="AT4g33150"
/number=4
intron complement (14256. .14351)
/gene="AT4g33150"
/number=4
exon complement (14352. .14583)
/gene="AT4g33150"
/number=5
intron complement (14584. .14664)
/gene="AT4g33150"
/number=5
exon complement (14665. .14823)
/gene="AT4g33150"
/number=6
intron complement (14824. .14919)
/gene="AT4g33150"
/number=6
exon complement (14920. .14999)
/gene="AT4g33150"
/number=7
intron complement (15000. .15100)
/gene="AT4g33150"
/number=7
exon complement (15101. .15230)
/gene="AT4g33150"
/number=8
intron complement (15231. .15353)
/gene="AT4g33150"
/number=8

Query Match 45.3%; Score 435; DB 8; Length 199536;

Best Local Similarity 86.3%; Pred. No. 9.8e-113; Mismatches 0; Indels 84; Gaps 1;

Matches 529; Conservative 0; 171 GTATTATGGAGGAACACTGTCTGAGAAAATGCTTGGTAAGGACCTAAAGGCTTTGCCAAGT 230
Db 199204 GTATTATGGAGGAACACTGTCTGAGAAAATGCTTGGTAAGGACCTAAAGGCTTTGCCAAGT 199145
Qy 231 CCTAGAAAGTGACTACATTTGGCTACTAAAGTGGTAGATATAAAGAGGTTTGGATTT 290
Db 199144 CCTAGAAAGTGACTACATTTGGCTACTAAAGTGGTAGATATAAAGAGGTTTGGATTT 199085
Qy 291 CAGTCTGAGAGAGTAAGAAAGAGTATTGACGAGAGCTTGGAGAGGCTTCAGCTTGATTA 350
Db 199084 CAGTCTGAGAGAGTAAGAAAGAGTATTGACGAGAGCTTGGAGAGGCTTCAGCTTGATTA 199025
Qy 351 TGTTCACATACCTTCATTGCCATGACATTTAGTTCCGGTCTCTTGAT----- 396
Db 199024 TGTTCACATACCTTCATTGCCATGACATTTAGTTCCGGTCTCTTGAT----- 198965
Qy 397 ----- 396
Db 198964 TATTTCCTCTGGGTTTATCACCTTTCGTTGTTGTGTCTCAAACTTTTACTAAGTGTTCAT 198905
Qy 397 -----CAGATTGTGAGTGAACAAATTCCTGCTCTTCAGAACTTCAACAGAGGG 446
Db 198904 TCATTGCTGCCAGATTGTGAGTGAACAAATTCCTGCTCTTCAGAACTTCAACAGAGGG 198845

Qy 447 GAAGACCCGGTTTCATTGGTATACATGGTCTTCCGTTAGATATTTTCACTTATGTTCTTGA 506
Db 198844 GAAGACCCGGTTTCATTGGTATACATGGTCTTCCGTTAGATATTTTCACTTATGTTCTTGA 198785
Qy 507 TCGAGTGCTCCAGGACTGTGATATTTGTCATACCTGTCATTACGGGCTTAAATGA 566
Db 198784 TCGAGTGCTCCAGGACTGTGATATTTGTCATACCTGTCATTACGGGCTTAAATGA 198725
Qy 567 TTTCGACGTTGCTGGATTTTACTACCTTACTTGAAGAGCAAGGTTGGGTGTGATTAAGTGC 626
Db 198724 TTTCGACGTTGCTGGATTTTACTACCTTACTTGAAGAGCAAGGTTGGGTGTGATTAAGTGC 198665
Qy 627 TTCTCCATTAGCAATGGGCTCTCTTACAGAACAAAGGTTCTCTGAAATGGCACCCCTGCTTC 686
Db 198664 TTCTCCATTAGCAATGGGCTCTCTTACAGAACAAAGGTTCTCTGAAATGGCACCCCTGCTTC 198605
Qy 687 CCCTGAGCTCAAG 699
Db 198604 CCCTGAGCTCAAG 198592

Search completed: November 13, 2005, 07:01:20
Job time : 4383 secs

THIS PAGE BLANK (NPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 13, 2005, 08:37:24 ; Search time 4913 Seconds
(without alignments)
3146.186 Million cell updates/sec

Title: US-10-606-300-11
Perfect score: 1619
Sequence: 1 MTKIELRALGNTGLKVSAG.....VEAILEPVKNLTWPSGIHQN 319

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DB=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO spool_h/US10606300/runat_07112005_094938_8177/app_query.fasta_1.455
-DB=GenEmbl -QFMT=fastcap -SURFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10606300@cgn 1 1 3731 @runat_07112005_094938_8177 -NCPU=6 -ICPU=3
-NO MWAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THRSADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sv:*
13: gb.un:*
14: gb.vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1619	100.0	960	6	AR405970 Sequence
2	1619	100.0	960	6	AX259907 Sequence
3	1619	100.0	960	6	AX374959 Sequence
4	1619	100.0	960	8	AJ417563 Arabidops

5	1619	100.0	960	8	AY090337	Arabidops
6	1619	100.0	1092	8	AY050377	Arabidops
c 7	1402.5	86.6	98124	8	ATT1611	AL031394 Arabidops
c 8	1402.5	86.6	192861	8	ATCHRIV80	AL161584 Arabidops
c 9	1402.5	86.6	192536	8	ATCHRIV79	AL161583 Arabidops
10	1348	83.3	1215	8	AY176585	Actinidia
11	1322	81.7	1221	8	AY264803	AY264803 Malus x d
12	1261	77.9	1239	8	AB160990	Spinacia
13	1215	75.0	1234	8	AK060431	Oryza sat
14	1215	75.0	1305	8	AK067039	Oryza sat
15	1211	74.8	1275	8	AK102223	Oryza sat
c 16	944.5	58.3	118803	8	AC124214	Medicago
c 17	912.5	56.4	142246	8	CNS08C88	AL731787 Oryza sat
18	697	43.1	303446	1	AE016941	Bacteroid
19	693.5	42.8	110000	2	AP006493	Continuation (6 of
20	668	41.3	2105	3	AK112151	AK112151 Ciona int
c 21	625	38.6	311050	1	BX294133	BX294133 Pirellula
22	593.5	36.7	1038	6	CQ600264	Sequence
c 23	546.5	33.8	3724	6	CQ600263	Sequence
c 24	546.5	33.8	20956	2	AC018265	AC018265 Drosophil
c 25	546.5	33.8	182182	3	AC007889	AC007889 Drosophil
c 26	546.5	33.8	226076	3	AE003694	AE003694 Drosophil
27	544.5	33.6	1051	6	CQ612720	Sequence
28	544.5	33.6	1191	3	AY118797	AY118797 Drosophil
c 29	517	31.9	3308	6	CQ612719	Sequence
c 30	458	28.3	3014	6	CQ579386	Sequence
c 31	457.5	28.3	3349	2	AC020400	AC020400 Drosophil
32	426	26.3	10901	1	AE012069	AE012069 Xanthomon
33	420.5	26.0	885	6	CQ579387	Sequence
34	417.5	25.8	11874	1	AE012528	AE012528 Xanthomon
c 35	414.5	25.6	110000	8	CR380952_05	Continuation (6 of
c 36	414.5	25.6	110000	8	CR380952_06	Continuation (7 of
c 37	413	25.5	110000	8	CR382132_38	Continuation (39 o
c 38	389.5	24.1	299300	1	AP005026	AP005026 Streptomy
39	387.5	23.9	110000	8	AE016815	Continuation (6 of
c 40	378	23.3	360	11	CR384305	CR384305 Arabidops
c 41	375.5	23.2	110000	8	CR382126_24	Continuation (25 o
c 42	374	23.1	33324	3	U00033	U00033 Caenorhabdi
43	368.5	22.8	299425	1	AP005049	AP005049 Streptomy
44	366	22.6	987	6	E07105	E07105 gDNA encodi
45	366	22.6	1368	1	PSEFUDH	D32042 Pseudomonas

ALIGNMENTS

RESULT 1
AR405970
LOCUS AR405970 960 bp DNA
DEFINITION Sequence 12 from patent US 6630330.
ACCESSION AR405970
VERSION AR405970.1 GI:40155041
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 960)
AUTHORS Porro,D. and Sauer,M.
TITLE Ascorbic acid production from yeast
JOURNAL Patent: US 6630330-A 12 07-OCT-2003;
FEATURES Location/Qualifiers
source
1..960
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 5.57e-130 Length: 960
Score: 1619.00 Matches: 319
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-606-300-11 (1-319) x AR405970 (1-960)

QY 1 MetThrLysIleGluLeuArgAlaLeuGlyAenThrGlyLeuLysValSerAlaValGly 20
Db 1 ATGACGAAATAGAGCTTCGAGCTTTGGGAAACACAGGCTTAAGGTTAGCGCGCTTGGT 60
QY 21 PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAla 40
Db 61 TTGTGGCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGAAGATGATCGCTCGCC 120
QY 41 ThrValArgGluAlaPheArgLeuGlyIleAenPheAspThrSerProTyrTyrGly 60
Db 121 ACCGTGGCGAGGCTTTCGCTCTCGGTATCACTTCTTCGACACCTCCCGTATTATGGA 180
QY 61 GlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSer 80
Db 181 GGAAACACTGCTCGAGAAATGCTTGGTAAGGACTAAAGGCTTTCGAAGTCCCTAGAAGT 240
QY 81 AspTyrIleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGlu 100
Db 241 GACTACATTGTGGCTACTAAGTGTGGTAGATATAAGAGGTTTTCGATTTCAGTCTGAG 300
QY 101 ArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 120
Db 301 AGAGTAAGAAAGAGTATTGACGAGAGCTTGGAGAGCTTCAGCTTGATTATGTTGACATA 360
QY 121 LeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePro 140
Db 361 CTTCAATTGCGCATGACATTGAGTTCGGTCTCTTGATCAGATTGTGAGTGAAACAATTCT 420
QY 141 AlaLeuGlnLysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGlyLeuPro 160
Db 421 GCTCTTCAGAAACTGAACAGAGGGAGACCCGGTTCAATGGTATCACTGGTCTTCGG 480
QY 161 LeuAspIlePheThrTyrValLeuAspArgValProGlyThrValAspValIleLeu 180
Db 481 TTAGATATTTTCACTTATGTTCTTGATCGAGTGCCTCCAGGACTGTGATGTGATATTG 540
QY 181 SerTyrCysHisTyrGlyValAenAspSerThrLeuLeuAspLeuLeuProTyrLeuLys 200
Db 541 TCATACATGTCATTACGGCGTTAATGATTCAGCTTGTGATTTACTACCTTACTTGAAG 600
QY 201 SerLysGlyValGlyValIleSerIleAspProLeuAlaMetGlyLeuLeuThrGluGln 220
Db 601 AGCAAGGTGTGGGTGTGATAAGTCTCTCCATTAGCAATGGGCTCTCTACAGAACAA 660
QY 221 GlyProGluTyrHisProAlaSerProGluLeuLysSerAlaSerLysAlaVal 240
Db 661 GGTCTCTGTAATGGCACCTGCTTCCCTCGAGCTCAAGTCTGCAAGCAAGCCGAGTT 720
QY 241 AlaHisCysLysSerLysGlyLysIleThrLysLeuAlaLeuGlnTyrSerLeuAla 260
Db 721 GCTCACTGCAATCAAGGCGCAAGAGATCAAAAGTTAGCTCTGCAATACAGTTTAGCA 780
QY 261 AsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGluAsn 280
Db 781 AACAAAGGAGATTTCGTGGTGTGGTGGGATGAGCTCTGTCTCAGAGTAGAAGAAAAT 840
QY 281 ValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluVal 300
Db 841 GTTGACAGATTACAGAGCTTGAAAGCTTGGGGATGGATCAAGAACTCTGTCTGAGTT 900
QY 301 GluAlaIleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGlnAen 319
Db 901 GAAGCTATTCTCGAGCTGTAAAGAATCTGACATGCCAAGTGAATCCATCAGAAC 957

RESULT 2

AX259907
LOCUS AX259907 960 bp DNA linear PAT 26-OCT-2001
DEFINITION Sequence 4 from Patent WO0172974.
ACCESSION AX259907
VERSION AX259907.1 GI:16508970
KEYWORDS

SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
REFERENCE Smirnoff, N.S. and Wheeler, G.L.
AUTHORS L-galactose dehydrogenase from Arabidopsis thaliana, and uses
TITILE thereof
JOURNAL Patent: WO 0172974-A 4 04-OCT-2001;
AScorbex Limited (GB)
FEATURES Location/Qualifiers
source 1..960
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
CDS 1..960
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD10214.1"
/db_xref="GI:16508971"
/translation="MTKIELRALGNLTGLKVSAGVFGSPGSLGVSFGVPAEDDAVATVRE
APRLGINFDTSPYVGTSLSEKMLGKALQVPRSDYIVATKCRYEKGFDSAEV
RKSIDESLERIQLDVDILHCHDIEFGSLDOISETIPALOKLQKQETRTFTIGLP
LDITYVLDRVPPGTVDVILSYCHGVNDSLLDLLPYLKSQGVISASPLAMGLLT
EQGPPEHPASPFLKSAKAAVAHCKSKGKIKTLAQISLANKEISSVLVGMSSVSQ
VEENVAATELESIGMDQETLSEVEAILPEPVKNTWPSGIHQN"
ORIGIN
Alignment Scores:
Pred. No.: 5,57e-130 Length: 960
Score: 1619.00 Matches: 319
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-606-300-11 (1-319) x AX259907 (1-960)
QY 1 MetThrLysIleGluLeuArgAlaLeuGlyAenThrGlyLeuLysValSerAlaValGly 20
Db 1 ATGACGAAATAGAGCTTCGAGCTTTGGGAAACACAGGCTTAAGGTTAGCGCGCTTGGT 60
QY 21 PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAla 40
Db 61 TTGTGGCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGAAGATGATCGCTCGCC 120
QY 41 ThrValArgGluAlaPheArgLeuGlyIleAenPheAspThrSerProTyrTyrGly 60
Db 121 ACCGTGGCGAGGCTTTCGCTCTCGGTATCACTTCTTCGACACCTCCCGTATTATGGA 180
QY 61 GlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSer 80
Db 181 GGAAACACTGCTCGAGAAATGCTTGGTAAGGACTAAAGGCTTTCGAAGTCCCTAGAAGT 240
QY 81 AspTyrIleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGlu 100
Db 241 GACTACATTGTGGCTACTAAGTGTGGTAGATATAAGAGGTTTTCGATTTCAGTCTGAG 300
QY 101 ArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 120
Db 301 AGAGTAAGAAAGAGTATTGACGAGAGCTTGGAGAGCTTCAGCTTGATTATGTTGACATA 360
QY 121 LeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePro 140
Db 361 CTTCAATTGCGCATGACATTGAGTTCGGTCTCTTGATCAGATTGTGAGTGAAACAATTCT 420
QY 141 AlaLeuGlnLysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGlyLeuPro 160
Db 421 GCTCTTCAGAAACTGAACAGAGGGAGACCCGGTTCAATGGTATCACTGGTCTTCGG 480
QY 161 LeuAspIlePheThrTyrValLeuAspArgValProGlyThrValAspValIleLeu 180
Db 481 TTAGATATTTTCACTTATGTTCTTGATCGAGTGCCTCCAGGACTGTGATGTGATATTG 540
QY 181 SerTyrCysHisTyrGlyValAenAspSerThrLeuLeuAspLeuLeuProTyrLeuLys 200
Db 541 TCATACATGTCATTACGGCGTTAATGATTCAGCTTGTGATTTACTACCTTACTTGAAG 600
QY 201 SerLysGlyValGlyValIleSerIleAspProLeuAlaMetGlyLeuLeuThrGluGln 220
Db 601 AGCAAGGTGTGGGTGTGATAAGTCTCTCCATTAGCAATGGGCTCTCTACAGAACAA 660
QY 221 GlyProGluTyrHisProAlaSerProGluLeuLysSerAlaSerLysAlaVal 240
Db 661 GGTCTCTGTAATGGCACCTGCTTCCCTCGAGCTCAAGTCTGCAAGCAAGCCGAGTT 720
QY 241 AlaHisCysLysSerLysGlyLysIleThrLysLeuAlaLeuGlnTyrSerLeuAla 260
Db 721 GCTCACTGCAATCAAGGCGCAAGAGATCAAAAGTTAGCTCTGCAATACAGTTTAGCA 780
QY 261 AsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGluAsn 280
Db 781 AACAAAGGAGATTTCGTGGTGTGGTGGGATGAGCTCTGTCTCAGAGTAGAAGAAAAT 840
QY 281 ValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluVal 300
Db 841 GTTGACAGATTACAGAGCTTGAAAGCTTGGGGATGGATCAAGAACTCTGTCTGAGTT 900
QY 301 GluAlaIleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGlnAen 319
Db 901 GAAGCTATTCTCGAGCTGTAAAGAATCTGACATGCCAAGTGAATCCATCAGAAC 957

JOURNAL Plant J. 30 (5), 541-553 (2002)
 MEDLINE 22043773
 PUBMED 12047629
 REFERENCE 2 (bases 1 to 960)
 AUTHORS Gatzek S.
 TITLE Direct Submission
 JOURNAL Submitted (23-OCT-2001) Gatzek S., School of Biological Sciences,
 University of Exeter, Exeter, EX4 4PS, UNITED KINGDOM
 FEATURES Location/Qualifiers
 1..960
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 1..960
 /gene="L-GALDH"
 /gene="L-GalDH"
 /function="oxidises L-galactose at C1, forming
 L-galactono-1,4-lactone"
 /codon_start=1
 /evidence=experimental
 /product="L-galactose dehydrogenase"
 /protein_id="CAD10386.1"
 /db_xref="GI:16555790"
 /db_xref="UniProt/TREMBL:O81884"
 /translation="MTKIELRALGNTGLKVSANGFASPLGVSFGFVADDDAVATVRE
 AFRGINFDTSPYGGTSLERKMLGKALQVPRSDYIVATKQYKEGDFDSAEVY
 RKSIDLERLDVDVILHCHDIFGLSDQIVSETIPALQKQEGKTRFRTIGLPL
 LDIFYDRDPVPGVTVILSYCHYGVNDSTLLDLLPYLKSQGVIVISASPLAMGLLT
 EOGPEHPHAPSLKASAKAAVAHCKSKGKTKLALQYSLANKEISSVLVGMSSVSQ
 VEENVAAVTELSGLMDQETLSEVALEPYNKLTWPSGIHQN"

source

gene

CDS

ORIGIN

Alignment Scores:
 Pred. No.: 5,57e-130 Length: 960
 Score: 1619.00 Matches: 319
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 8 Gaps: 0

US-10-606-300-11 (1-319) x ATH17563 (1-960)

QY 1 MetThrLysIleGluLeuArgAlaLeuGlyAenThrGlyLeuLysValSerAlaValGly 20
 DB 1 ATGACGAAATAGAGCTTCGAGCTTTGGGGAACACAGGCGCTTAAGGTTAGCGCCGTGCT 60
 QY 21 PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAla 40
 DB 61 TTGGTGCCCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCCGCAAGATGATCCGCTGCC 120
 QY 41 ThrValArgGluAlaPheArgLeuGlyIleAenPheAspThrSerProTyrTyrGly 60
 DB 121 ACCGTGCGCGAGGCTTTCGCTCGGTATCAACTTCTTCGACACCTCCCGTATTATGCA 180
 QY 61 GlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSer 80
 DB 181 GGAACACTCTCTGAGAAATGCTTGGTAAGGGAGCTAAAGGCTTTCGAGCTCCCTAGAAGT 240
 QY 81 AspTyrIleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGlu 100
 DB 241 GACTACATTGTGCTACTAAGTGTGTGTAGATATTAAGAAGGTTTGTATTTCAGTCTGAG 300
 QY 101 ArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 120
 DB 301 AGAGTAAGAAGAGTATTGACGAGAGCTTGGAGAGGCTTCAGCTTATTATGTTGACATA 360
 QY 121 LeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePro 140
 DB 361 CTTCAATGTCATGACATTGAGTTCCGGGTCTCTTGATCAGATTGTGAGTGAACAAATTCCT 420
 QY 141 AlaLeuGlnLysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGlyLeuPro 160

Db 421 GCTCTTCAGAACTGAACAAGAGGGGAGACCGGTTTCATTGGTATCACTGGTCTTCG 480
 QY 161 LeuAspIlePheThrTyrValLeuAspArgValProGlyThrValAspValIleLeu 180
 DB 481 TTAGATAATTTTCACTTATGTTCTTGATCGAGTCCCTCCAGGAGCTGTCTGATGATTTG 540
 QY 181 SerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLys 200
 DB 541 TCATCTGTCATTACGGCGTTAATGATTGACGTTGCTGGATTTACTACCTTACTTTGAG 600
 QY 201 SerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGln 220
 DB 601 AGCAAGGTGTGGTGTGATTAAGTCTTCTCCATTAGCAATGGGCTCTCTTACAGACAA 660
 QY 221 GlyProGluTyrHisProAlaSerProGluLeuLysSerAlaSerLysAlaAlaVal 240
 DB 661 GGTCTCTCTGAATGGCACCTGCTTCCCTCAGCTCAAGTCTGCAAGCAAGCCGAGTT 720
 QY 241 AlaHisCysLysSerLysGlyLysIleThrLysLeuAlaLeuGlnTyrSerLeuAla 260
 DB 721 GCTCAGTCAAAATCAAGAGGCAAGATCAAGATTAGCTCTGCAATACAGTTAGCA 780
 QY 261 AsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGln 280
 DB 781 AACAGGAGATTTCGTCGGTGTGGTGGATGAGCTCTGCTCACAGTAGAAGAAAT 840
 QY 281 ValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluVal 300
 DB 841 GTTGACGAGTTACAGAGCTTCAAAAGTCTGGGATGGATCAAGAACTCTGTCTGAGTT 900
 QY 301 GluAlaIleLeuGluProValLysAsnLeuThrTyrProSerGlyIleHisGlnAsn 319
 DB 901 GAAGCTATTCTCGAGCCTGTAAAGAAATCTGACATGCGCAAGTGGAAATCCATCAGAAC 957

RESULT 5
 AY090337 960 bp mRNA linear PLN 24-MAR-2002
 Arabidopsis thaliana AT4933670/T16L1_160 mRNA, complete cds.

DEFINITION AY090337
 ACCESSION AY090337
 VERSION AY090337.1 GI:19699263
 KEYWORDS FLI CDNA.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 960)

AUTHORS Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,
 Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
 Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
 Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T.,
 Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C.,
 Yamada K., Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W.,
 Theologis A. and Ecker J.R.

Arabidopsis ORF clones
 Unpublished

2 (bases 1 to 960)

AUTHORS Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,
 Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
 Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
 Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T.,
 Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C.,
 Yamada K., Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W.,
 Theologis A. and Ecker J.R.

Direct Submission

TITLE Submitted (14-MAR-2002) Salk Institute Genomic Analysis Laboratory
 JOURNAL (SIGHU), Plant Biology Laboratory, The Salk Institute for
 Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
 USA

COMMENT RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL CDNA : 'RIKEN

Arabidopsis Full-Length cDNA') : Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGECC (SSP) Consortium members carried out the sequencing and annotation of the RAPL cDNAs: Cheuk, R., Chen, H., Kim, C.J., Meyers, M.C., Shinn, P., Ban, J., Bowser, L., Chan, M.W., Chang, E., Dale, J.M., Deng, J.M., Goldsmith, A.D., Jones, T., Karlin-Neumann, G., Lam, B., Lee, J.M., Lin, J., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Davis, R.W., Theologis, A., and Ecker, J.R.

Cheuk, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk) contributed equally to this work as PIs.

FEATURES

source

Location/Qualifiers

1..960

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/chromosome="4"

/clone="U13146"

/ecotype="Columbia"

/note="this clone is in PUN1 51"

1..960

/note="unknown protein"

/codon_start=1

/product="AT4g33670/tl6L1_160"

/protein_id="AAL90998.1"

/db_xref="GI:19699264"

/translation="MTKIELRALGNTGLKSVAGFGASPLGSVFPVABDDAVATVRE

APRLINFDSPYGGTSLSEKMLGKGLQVPSDYIVATKCRVKEGDFDFAERV

RKSIDLERLDVVDILHCHDIEFGLSDIOVRSITPALOKLQKQKTRFIGHTGLP

LDIFTYVLDVPPGTVDSLVCHYGVNDSLLDLLPYLKSQGVISASPLAMGLLT

EQPPEWHPASPFLKSKAAVAHCKSGKKITKLQYSLANKEISSVLVGMSSVSQ

VEENVAATELESIGMDQETLSEVAILEPKNLTWPSGIHQN"

CDS

ORIGIN

Alignment Scores:
Pred. No.: 5, 57e-130 Length: 960
Score: 1619.00 Matches: 319
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0

US-10-606-300-11 (1-319) x AY090337 (1-960)

QY 1 MetThrLysIleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGly 20
DB 1 ATGACGAAATAGAGCTTCGAGCTTTGGGAAACACAGAGCTTAAGGTTAGCGCGTGGT 60
QY 21 PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAla 40
DB 61 TTTGGTGCCTCTCCGCTCGAAGTGTCTTCGCTCCAGTCGCGCGAAGATGATCGCGTGC 120
QY 41 ThrValArgGluAlaPheArgLeuGlyIleAsnPheAspThrSerProTyrTyrGly 60
DB 121 ACCGTGCGGAGGCTTTCGCTCGGTATCAACTTCTTCGACACCTCCCGGTATTATGGA 180
QY 61 GlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSer 80
DB 181 GGAACACTGTCTGAGAAATGCTTGGTAAGGACTAAAGCTTTCGAAGTCCCTAGAGT 240
QY 81 AspTyrIleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGlu 100
DB 241 GACTACATTGGCTACTAAGTGTGGTAGATATAAAGAGGTTTTCATTTCAGTCTCAG 300
QY 101 ArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 120
DB 301 AGAGTAAGAAGAGTATTGACGAGAGCTTGGAGAGCTTCAGCTGATTATGTTGACATA 360

QY 121 LeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePro 140
DB 361 CTTCAATGCCATGACATTGAGTTCGGGTCTCTTGTATCAGATTGTGAGTGAACAATTCCT 420
QY 141 AlaLeuGlnLysLeuLysGlnGluGlyThrArgPheIleGlyIleThrGlyLeuPro 160
DB 421 GCTCTTCAGAACTGAAACAAGAGGGAGACCGGTTTCATTTGGTATCCTGCTTCCG 480
QY 161 LeuAspIlePheThrTyrValLeuAspArgValProGlyThrValAspValIleLeu 180
DB 481 TTAGATATTTTCACTTATGTTCTTGTATGATGAGTCTCCAGGAGCTGTGATGTGATATTG 540
QY 181 SerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLys 200
DB 541 TCATCTGTCTATACGGCGTTAATGATTCGACGTTGCTGGGATTTACTACCTTACTTGAAG 600
QY 201 SerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuThrGluGln 220
DB 601 AGCAAGGTGTGGGTGTGATAGTCTTCTCATTTAGCAATGGGCTCTTACAGACAA 660
QY 221 GlyProGluTyrHisProAlaSerProGluLeuLysSerAlaSerLysAlaVal 240
DB 661 GGTCTCTCTGATGGCACCCTGCTTCCCTGAGCTCAAGTCTGCAAGCAAGCGCAGTT 720
QY 241 AlaHisCysLysSerLysGlyLysIleThrLysLeuAlaLeuGlnTyrSerLeuAla 260
DB 721 GCTCACTGCAATCAAGGGCAAGAGATCAAAAGTTAGCTCTGCAATACAGATTAGCA 780
QY 261 AsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGluAsn 280
DB 781 AACAGGAGATTTCTGCTGGTGTGGTGGATGAGCTCTGCTCACAGGTAGAAGAAAT 840
QY 281 ValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluVal 300
DB 841 GTTCGACAGTTACAGAGCTTGAAGTCTGGGATGATCAAGAACTCTGCTCAGGTT 900
QY 301 GluAlaIleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGlnAsn 319
DB 901 GAAGTATTCTCGAGCCTGTAAGAATCTGACATGGCCCAAGTGGAAATCCATCAGAAC 957

RESULT 6
AY050377

LOCUS
DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana

Arabidopsis thaliana

REFERENCE

AUTHORS

1 (bases 1 to 1092)

Koesema, E., Chen, H., Cheuk, R., Kim, C.J., Meyers, M.C., Shinn, P.,

Bath, J., Bowser, L., Carninci, P., Dale, J.M., Goldsmith, A.D.,

Hayashizaki, Y., Ishida, J., Jiang, P.X., Jones, T., Kamiya, A.,

Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X.,

Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J.,

Pham, P.K., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A.,

Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S.,

Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.

Unpublished

2 (bases 1 to 1092)

Koesema, E., Chen, H., Cheuk, R., Kim, C.J., Meyers, M.C., Shinn, P.,

Bath, J., Bowser, L., Carninci, P., Dale, J.M., Goldsmith, A.D.,

Hayashizaki, Y., Ishida, J., Jiang, P.X., Jones, T., Kamiya, A.,

Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X.,

Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J.,

Pham, P.K., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A.,

Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S.,

Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.

Direct Submission

TITLE

JOURNAL

Submitted (06-AUG-2001) Salk Institute Genomic Analysis Laboratory (SIGNAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

COMMENT

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPL cDNAs (RAPL cDNA : 'RIKEN Arabidopsis Full-length cDNA') : Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGECC (SSP) Consortium members carried out the sequencing and annotation of the RAPL cDNAs: Koesema, E., Chen, H., Cheuk, R., Kim, C.J., Meyers, M.C., Shinn, P., Bani, J., Bowser, L., Dale, J.M., Goldsmith, A.D., Jiang, P.X., Jones, T., Karlin-Neumann, G., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K., Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Davis, R.W., Theologis, A., and Ecker, J.R.

Koesema, E. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk) contributed equally to this work as PIs.

FEATURES

Location/Qualifiers

1..1092

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/chromosome="4"

/clone="RAPL07-17-C15 (R13146)"

/ecotype="Columbia"

1..7

8..967

/note="unknown protein"

/codon_start=1

/product="AT4g33670/Ti16L1_160"

/protein_id="AAK91395.1"

/db_xref="GI:15215698"

/translation="MTKIELRALGNTGLKVSANGFGASPLGSVFGFVAEDDAVATVRE
AFRLGNFDPYGGTTSERMLGKGLKALQVPSDYIVATKCGRYKEGDFSAERV
RKSIDESLERLDQVDILHCDIEFGSLDQIVSETIPALQKQEGKTRFIIITGLP
LDIFTVLDRVPDVTDLISYCHGVNDTLDLLPYLKSNGVGVISASPLWMLLT
EGPEWHPASPDELKASAKVAHCKSKGKITKLALQYSLANKEISSVLVGMSSVSQ
VEENVAATELESGLGMDQTLSEVEALLEPKNLTPSGIHQN"

968..1092

3' UTR

ORIGIN

Alignment Scores:

Pred. No.: 6,58e-130 Length: 1092
Score: 1619.00 Matches: 319
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0

US-10-606-300-11 (1-319) x AY050377 (1-1092)

QY 1 MetThrLysLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGly 20
DB 8 ATGACGAAATAGAGCTTCGAGCTTTGGGGAACACACAGGCTTAAGGTTAGCGCCGTTGGT 67
QY 21 PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAla 40
DB 68 TTTGGTGCCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGAAGATGATCCGCTCGCC 127
QY 41 ThrValArgGluAlaPheArgLeuGlyIleAsnPheAspThrSerProTyrTyrGly 60
DB 128 ACGTCGGCGAGCTTTCGTCGGTATCAACTTTCGACACCTCCCGATTATGGA 187
QY 61 GlyThrLeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSer 80
DB 188 GGAACACTGCTCAGAAAATGCTTGTGTAAGGAGCTAAAGGCTTTCGAACTCCCTAGAACT 247
QY 81 AspTyrIleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGlu 100

DB 248 GACTACATGTGGCTACTAAAGTGGTAGATATAAAGAAAGGTTTTTGATTTTCAGTCTGAG 307
QY 101 ArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 120
DB 308 AGAGTAAGAAAGATATTGACGAGAGCTTGCAGAGGCTTCAGCTTGATTATGTTGACATA 367
QY 121 LeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePro 140
DB 368 CTTCAATGCCATGACATTGAGTTCCGGTCTCTTGATCAGATTGTGAGTGAACAATTCCT 427
QY 141 AlaLeuGlnLysLeuLysGlnGluGlyThrArgPheIleGlyIleThrGlyLeuPro 160
DB 428 GCTCTTCAGAACTGAAACAAGAGGGGAAGACCCGGTTCATTGGTATCCTGCTTCCG 487
QY 161 LeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeu 180
DB 488 TTAGATATTTTCACTTATGTTCTTGATCGAGTGCCTCCAGGAGCTGTCGATGTGATATTG 547
QY 181 SerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrIleLys 200
DB 548 TCATACTGTCATTACGGCGTTAATGATTGCGAGCTTGTGCGATTACTACTCTTACTTGAAG 607
QY 201 SerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGln 220
DB 608 AGCAAGGTGGGGTGTGATAGTCTTCTCCATTGCAATGGGCTCTTACAGAACAA 667
QY 221 GlyProGluTyrHisProAlaSerProGluLeuLysSerAlaSerLysAlaVal 240
DB 668 GGTCTCTCTGATGCGACCTTCTCCCTGAGCTCAAGTCTGCAAGCAAGCCAGTT 727
QY 241 AlaHisCysLysSerLysGlyLysIleThrLysLeuAlaLeuGlnTyrSerLeuAla 260
DB 728 GCTCACTGCAAAATCAAGAGGCAAGATCAAAAGTGTAGCTCTGCAATACAGTTTAGCA 787
QY 261 AsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGluAsn 280
DB 788 AACAAAGAGATTTTCGTCGGTGTGGTGGATGAGCTCTGTCTCACAGTAGAAGAAAT 847
QY 281 ValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluVal 300
DB 848 GTTGCGAGCTTACAGAGCTTGAAAGTCTGGGATGGATCAAGAACTCTGTCTGAGGTT 907
QY 301 GluAlaIleLeuGluProValLysAsnLeuThrTyrProSerGlyIleHisGlnAsn 319
DB 908 GAAGCTATTCTCGAGCCTGTAAAGAAATCTGATGCGCAAGTGAATCCATCAGAAC 964
RESULT 7
ATT16L1/c 98124 bp DNA linear PLN 27-AUG-1999
LOCUS Arabidopsis thaliana DNA chromosome 4, BAC clone Ti6L1 (ESSAII project).
DEFINITION
ACCESSION AL031394
VERSION AL031394.1 GI:3859658
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE 1 Bevan, M., Obermaier, B., Deutschenbaur, S., Piravandi, E.,
AUTHORS Hohnel, J., Jesse, T., Heijnen, L., Vos, P., Mewes, H.W., Mayer, K.F.X.
and Schueller, C.
2 Unpublished
3 (bases 1 to 98124)
JOURNAL EU Arabidopsis sequencing project.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (09-NOV-1998) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk

COMMENT On Nov 11, 1998 this sequence version replaced gi:3549653.
FEATURES
Location/Qualifiers
1..98124
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
1..6620
/gene="T16L1.10"
CDS
join(<1..228,315..368,536..689,1974..2048,2138..2415,
2754..2981,3156..3254,3388..3537,3662..3808,4065..4115,
4197..4354,4506..4638,4724..4821,5206..5452,5701..5904,
6212..6292,6393..6620)
/gene="T16L1.10"
/note="Contains E1-E2 ATPases phosphorylation site
[DKTGLT], Heavy-metal-associated domain
[VGMTCGCGSASVKKILESQQVASASVNL]"
/codon_start=1
/product="metal-transporting P-type ATPase (fragment)"
/protein_id="CAA20565.1"
/db_xref="GI:3549654"
/db_xref="GOA:O9SZC9"
/db_xref="UniProt/Swiss-Prot:O9SZC9"
/translation="AVLPVIRHRLCLSSSPFRSISGGSGFGGNGSGGGGG
GSESGDSKLGANADGVSPSSDIIILDVGMTCGCGSASVKKILESQQVASASV
NLTTTALVWPPEAKSPDWKSLGETLANHLTNCGFQSTPRDLVTENPKFVETKT
KDQARLESRELAVALCAVCLVGHLLTFLGVNAPWIAHISTGFHSLCLITLL
GPKRLVDGIGKSLKSGPNMTLVGLCASFSVSLAAMI PKLGWKTFFPEPMLI
APVLLGRNLEQBAKIKATSDMTGLLSVLPKARLLLDGLONSTVEPCNSLSVGDV
VILPDVADGVKSGSTIDESSFTGEPLPVTKESGQVAAGSINLNGLTVEVHR
SGGETAVGDIIRLVEEAQSRAPVQQLVDKVGRTYGVMAASAATFTFWLFGAHL
PSALHNGPMSLQGLSCVLVACALGATPTAMLVGTSLGARRGLLRGGDILE
KPSLVDVTFKPTGLTKGHPVTVTEIIPENPRNLNDTSEVLEMLAAAVESNTH
PVGKAIKAARNCQTKAEDGTTEPBGSAVAIVNKRVTVTGLEWKRHGATGN
SLLALEEHNIVYIGVDNTLAARFEPKVEDAAQVVENLTROGIDVYMLSGD
KRNAAYVASVVGNIHVIAGVKAPEKKNFINELQKNKIVAMVGGDINDAALAS
NVGAMGGGAASVSPVLMGNRLTQLDAMELSRQTMKTIVKQNLWWAFGYNIVGI
PIAAGVLLPTGLTMTPTSMAGALMGVSLGVTNLSLLRYFRFSNRNDKNVKEPKEG
TKQPHENTRWKQSS"
1..228
/gene="T16L1.10"
/number=1
229..314
/gene="T16L1.10"
/number=1
315..368
/gene="T16L1.10"
/number=2
369..535
/gene="T16L1.10"
/number=2
536..689
/gene="T16L1.10"
/number=3
690..1973
/gene="T16L1.10"
/number=3
1974..2048
/gene="T16L1.10"
/number=4
2049..2137
/gene="T16L1.10"
/number=4
2138..2415
/gene="T16L1.10"
/number=5
2416..2753
/gene="T16L1.10"
/number=5
2754..2981
/gene="T16L1.10"
/number=6

2982..3155
/gene="T16L1.10"
/number=6
3156..3254
/gene="T16L1.10"
/number=7
3255..3387
/gene="T16L1.10"
/number=7
3388..3537
/gene="T16L1.10"
/number=8
3538..3661
/gene="T16L1.10"
/number=8
3662..3808
/gene="T16L1.10"
/number=9
3809..4064
/gene="T16L1.10"
/number=9
4065..4115
/gene="T16L1.10"
/number=10
4116..4196
/gene="T16L1.10"
/number=10
4197..4354
/gene="T16L1.10"
/number=11
4355..4505
/gene="T16L1.10"
/number=12
4506..4638
/gene="T16L1.10"
/number=12
4639..4723
/gene="T16L1.10"
/number=12
4724..4821
/gene="T16L1.10"
/number=13
4822..5205
/gene="T16L1.10"
/number=13
5206..5452
/gene="T16L1.10"
/number=14
5453..5700
/gene="T16L1.10"
/number=14
5701..5904
/gene="T16L1.10"
/number=15
5905..6211
/gene="T16L1.10"
/number=15
6212..6292
/gene="T16L1.10"
/number=16
6293..6392
/gene="T16L1.10"
/number=16
6393..6620
/gene="T16L1.10"
/number=17
7274..11125
/gene="T16L1.20"
complement(join(7274..7811,7900..8422,8677..8931,
9058..9166,9265..9317,9422..9682,10151..10429,
10510..10738,10832..11125))
/gene="T16L1.20"
complement(join(7274..7811,7900..8422,8677..8931,

CDS


```
Db 50660 ATTCAAGCTTGTTATTCCTGTCAGGTCCTCAAGTCCCAAAATATGTTGGGAAGTTGGAGTTAAATAT 50601
Qy 276 -----GlnValGluGluAenValAlaAlaValThrGluLeu 287
Db 50600 GAAAGAAATTTGTTCTGTTCTGAAACAGGTAGAGAAATGTTGCACAGCTTACAGAGCTT 50541
Qy 288 GluSerLeuGlyMetAspGlnGluThrLeuSerGluValGluAlaLeuGluProVal 307
Db 50540 GAAAGCTGGGATGCATCAAGAAACTCTGTGAGGTGAAGCTATTCTCGAGCCTGTA 50481
Qy 308 LysAenLeuThrTrpProSerGlyIleHisGlnAen 319
Db 50480 AAGAATCTGACATGGCCAAGTGAATCCATCATCAAC 50445

RESULT 8
ATCHRIV80/c 192861 bp DNA linear PLN 16-MAR-2000
LOCUS Arabidopsis thaliana DNA chromosome 4, contig fragment No. 80.
DEFINITION AL161584
ACCESSION AL161584.2 GI:7270316
VERSION
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE
1 (bases 1 to 50906)
Obermaier B., Deutschenbaur, S., Piravandi, E., Mewes, H.W., Lemcke, K.
and Mayer, K.P.X.
Unpublished
JOURNAL
REFERENCE
2 (bases 49025 to 135772)
Vitale, D., Liguori, R., Argiriou, A., De Simone, V., Mewes, H.W.,
Lemcke, K. and Mayer, K.P.X.
Unpublished
JOURNAL
REFERENCE
3 (bases 130050 to 192861)
Weichselgartner, M., Fartmann, B., Granderath, K., Dauner, D.,
Herzl, A., Neumann, S., Mewes, H.W., Lemcke, K. and Mayer, K.P.X.
Unpublished
JOURNAL
REFERENCE
4 (bases 1 to 192861)
EU Arabidopsis sequencing, project.
Direct Submission
Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
COMMENT
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
this fragment has an overlap with ATCHRIV79 at the 5' end and an
overlap with ATCHRIV81 at the 3' end.
FEATURES
Location/Qualifiers
1..192861
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosomes="4"
5401..7841
/gene="AT4g33680"
complement(join(5401..5694,5785..5991,6067..6387,
6489..6539,6625..6729,6942..7004,7289..7405,7601..7668,
7787..7841))
/gene="AT4g33680"
complement(join(5401..5694,5785..5991,6067..6387,
6489..6539,6625..6729,6942..7004,7289..7405,7601..7668,
7787..7841))
/gene="AT4g33680"
/note="similarity to aminotransferase (AspC family),
Aquifex aeolicus, PIR2:D70479
Contains Aminotransferases class-I pyridoxal-phosphate
source
gene
gene
CDS
attachment site AA267-280;Prokaryotic membrane lipoprotein
lipid attachment site AA312-322"
/codon_start=1
/product="putative protein"
/translacion="MAKRVNTCKCVATPQEKIEYKTVKSRNSNMKLOAGYLPPEIAR
RRSAHLKYDPAQVIVSLIGDITTEPIDEVITTSAMAKKAHELSTIEGSGYGAEOGAKP
LRAAIATFYGLGIGDDDFVSDGAKCDISRLQVMFGSNVTIAVDPSYPAYVDSVV
IMGOTGFNDVOKYKLEYMRCCTPENGFPEDLSTVGRTDIIIFCSNNPTGAAATRE
QLTQVFEAKNGSIIIVDSYANMYMSDDNPRSIFEIPGAEEVAMETASFSKYAGFTG
VRLGKENTNIIIDTFTSLGYDVGGKNAPYVTVVWHFPNQSSWDVFAILEKTHVVTPTG
SGFGGEGFVRVSAFGHRENIIEACRRFKQLYK"
/gene="AT4g33680"
complement(5401..5694)
/number=1
complement(5785..5991)
/number=2
complement(5991..6066)
/number=2
complement(6067..6387)
/number=3
complement(6388..6488)
/number=3
complement(6489..6539)
/number=4
complement(6540..6624)
/number=4
complement(6625..6729)
/number=5
complement(6730..6941)
/number=5
complement(6942..7004)
/number=6
complement(7005..7288)
/number=6
complement(7289..7405)
/number=7
complement(7406..7600)
/number=7
complement(7601..7668)
/number=8
complement(7669..7786)
/number=8
complement(7787..7841)
/number=9
8752..9691
/gene="AT4g33690"
join(8752..8920,9015..9691)
/gene="AT4g33690"
/codon_start=1
/product="hypothetical protein"
```

```
/protein_id="CAB80086.1"
/db_xref="GI:7270318"
/translation="MDLETENRIASILLREAAELRRQAEKQGVAYLEKPNVRRHPNS
RLTATVLGVQANAVETNEMWSLRSEIEFGERLKRKRDSSENGSQEONRRDFL
KRCTSVDENVTTISLSPSSRSRKRQSEDDDDGLGDVEVKTFLQSRHYHTQMLP
DYFLNNDRIWLLILKVLIVYDRVKRGRSGVGMMDPELFCPLPERELSKTSITGD
RKLVTPERSPLLRRTDSSSDEBEVYKRAHRKRKEHKKKLKKHKSKKKRDRKR
KYGRD"
exon      8752..8920
           /gene="AT4g33690"
           /number=1
intron    8921..9014
           /gene="AT4g33690"
           /number=1
exon      9015..9691
           /gene="AT4g33690"
           /number=2
gene       10101..12658
           /gene="AT4g33700"
           complement(join(10101..10145,10220..10491,10571..10686,
10760..10833,10960..11047,11157..11296,11383..11469,
11562..11675,11811..11939,11982..12058,12565..12658))
           /gene="AT4g33700"
           complement(join(10101..10145,10220..10491,10571..10686,
10760..10833,10960..11047,11157..11296,11383..11469,
11562..11675,11811..11939,11982..12058,12565..12658))
           /note="similarity to hemolysin, 38k, Synechocystis sp.,
Pir2.S75595
contains EST gb:AT999192.1"
           /codon_start=1
           /product="putative protein"
           /protein_id="CAB80087.1"
           /db_xref="GI:7270319"
           /db_xref="UniProt/TREMBL:C81887"
           /translation="MSGLTGLMSLSLVLEVLAKSGTPHKKYAAKILPVVQNHL
LVTLLICNAAMEVSGMVFDDSDLSFQPLIFLDGLTAWGAAILISVTLILLFGEII
POISICRYGLAIGATVAFVFLVFCIPVAMPISKLJDLFLGHRRAALFRAELKTL
VDFHNGAEKGELTHDETTIAGALELSEKMKDAMPISDFVIDINAKDRDLMN
LILEKHSRVPVYOPTNIIGLVLKLLTINPDEEIPVKNTVIRRIPEVPEILPY
DLNRFQKLSHMAVVVRQCDKIHLPLSKNGSVKEARVDVDEGTPTPQERMLPKRS
LQKWSFPNRASSFKGSGSKSKKNDADILQLNGNPLPKLASBEAAVGIIITMEDVI
ESLLQEEIFDETDDHFEDS"
exon      complement(10101..10145)
           /gene="AT4g33700"
           /number=1
intron    complement(10146..10219)
           /gene="AT4g33700"
           /number=1
exon      complement(10220..10491)
           /gene="AT4g33700"
           /number=2
intron    complement(10492..10570)
           /gene="AT4g33700"
           /number=2
exon      complement(10571..10686)
           /gene="AT4g33700"
           /number=3
intron    complement(10687..10759)
           /gene="AT4g33700"
           /number=3
exon      complement(10760..10833)
           /gene="AT4g33700"
           /number=4
intron    complement(10834..10959)
           /gene="AT4g33700"
           /number=4
exon      complement(10960..11047)
           /gene="AT4g33700"
           /number=5
```

Pred. No.:	2.17e-108	Length:	192861
Score:	1402.50	Matches:	317
Percent Similarity:	53.55%	Conservative:	0
Best Local Similarity:	53.55%	Mismatches:	2
Query Match:	86.63%	Indels:	275
DB:	8	Gaps:	4
US-10-606-300-11 (1-319) x ATCHRIV80 (1-192861)			
Qy	1	MetThrLysIleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGly	20
Db	5000	ATGACGAAAATAGAGCTTCGAGCTTTGGGGAACACAGGGCTTAAGTTAGCGCGTGGT	4941
Qy	21	PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAla	40
Db	4940	TTTGGTGCTCTCCGCTCGGAAGTGTCTTCGGTCCAGTCGCGCAAGATGATGCGCGCC	4881
Qy	41	ThrValArgGluAlaPheArgLeuGlyIleAsnPhePheAspThrSerPro	57
Db	4880	ACCGTGCAGGAGCTTCCTCGTATCAACTTCCTTCGACACCTCCCGTAAGCATT	4821
Qy	57	-----	57
Db	4820	CTTCTCTTTATATTTCTAAAGATTACTGATTCTTCAATACTAATGCTGATTCTCTG	4761
Qy	57	-----	57
Db	4760	AATGTTGAGACTAATTAACTGTTAAAGTCAAAATTTTATTGGAGGTTTCATATTTAC	4701
Qy	58	-----	66
Db	4700	TTTTTTATTTGGCATTTATTGTGTGTGAAGGTATTATGAGGAACACTGTCTGAGAA	4641
Qy	67	MetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAspTyrIleValAlaThr	86
Db	4640	ATGCTTGTAGGAGACTAAGGCTTTCAGTCCCTAGAGTGACTACATGTTGCTACT	4581
Qy	87	LysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGluArgValArgLysSerIle	106
Db	4580	AAGTGTGTAGATATAAAGAGGTTTGTATTTCACTGCTGAGAGAGTAGAAGAGTAGT	4521
Qy	107	AspGluSerLeuGluArgLeuGlnLeuAspTyrValAlaSerIleLeuHisCysHisAspIle	126
Db	4520	GACGAGAGCTTGGAGAGGCTTCAGCTTATGTTGATGATACACTTCATTCCTGACATT	4461
Qy	127	GluPheGlySerLeuAsp	132
Db	4460	GAGTTGCGGCTCTCTTGATCAGGTTTACCACCTTATTCTCTTGGGTTTATCATCTTTCGTT	4401
Qy	133	-----	138
Db	4400	GTTTGTGTCTCAACCTTTTACTAAGTGTTCATTCATTGCTGCGCAGATTGTGAGTGAACA	4341
Qy	139	IleProAlaLeuGlnLysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGly	158
Db	4340	ATTCTGCTCTTCAGAAACTGAAACAGAGGGAGACCCCGGTTTCATTGTTACTGTTGT	4281
Qy	159	LeuProLeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAspVal	178
Db	4280	CTTCCGTTAGATATTTTCACTTATGTTCTTGATCGAGTCCCTCCAGGAGCTGCTCGATGG	4221
Qy	179	IleLeuSerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyr	198
Db	4220	ATATTGTTCATCTGTTCATTACGGCGTTAATGATTCGACGTTGCTCGATTTTACTACTTAC	4161
Qy	199	LeuLysSerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThr	218
Db	4160	TTGAAGAGCAAGGTTGGGTGTGTAAGTGTCTTCCATTAGCAATGGGCTCTCTTACA	4101
Qy	219	GluGlnGlyProProGluTrpHisProAlaSerProGluLeuLysSer	234
Db	4100	GAACAAGGTCTCTCTGAATGGCACCTCTGCTTCCCTCGAGTCAAGGT-TTGGCATTTTAC	4042

```
QY 234 ----- 234
Db 4041 AAAAGAATGTGATTTCTTTAAACATTTCTGTAGCACTTTGTATTGATACACCGTATAAGAG 3982
QY 234 ----- 234
Db 3981 ACTAGTGTAAACGCCATTCGTGAGCATTCCTCGAGCATAGGAGGTTAAATTTCTAT 3922
QY 234 ----- 234
Db 3921 AGATGCTATACCTCTGAGACTGACAATCTTTGTTCTTTAACTCGPACAATCATTTGTGATT 3862
QY 234 ----- 234
Db 3861 AGTCTTCATAGGCAAGATTAGAGACATTCAGATAGTTCCCAATGGGTATTAAACATGTTA 3802
QY 234 ----- 234
Db 3801 GTTGTCTTTGATGATTAATTCGTAGAGATTAAGTAAACAGATTCACATGACGCTGTTATTT 3742
QY 234 ----- 234
Db 3741 AGTTGAATTTGATTTTGAAGAGTAGGGATGCTCTTGAGTTAGTTTGGTTTGTCTCT 3682
QY 234 ----- 234
Db 3681 GTTGTATATTTGATTTAGTGAGAGTAAACGAGATACTGACTTCTTTAAACAAATTTGTTTC 3622
QY 235 ----- 237
Db 3621 TCAATATGAAGTCAATACCGTTGCTGCTTTTATTTCTTATGATGCTGCAAGCAAA 3562
QY 238 AlaAlaValAlaHisCysLeuSerLysGlyLysLysIleThrLysLeuAlaLeuGlnTyr 257
Db 3561 GCCGAGTTGCTCACTGCATCAATCAAGGCGAAGATCACAAAGTTAGCTCTGCAATAC 3502
QY 258 SerLeuAlaAsnLysGluIleSerSerValLeuValGlyMetSerSerValSer----- 275
Db 3501 AGTTTAGCAAAACAGGAGATTTGCTCGGTGTTGGTTGGGATGAGCTCTGTCTC-ACAGGT 3443
QY 275 ----- 275
Db 3442 ATTCAAGCTTGATTTCTCTGTCAGTCCCAATATATGTGGGAAGTTGGAGTTAAATAT 3383
QY 276 -----GlnValGluGluAsnValAlaAlaValThrGluLeu 287
Db 3382 GAAAGAAATTTGTGTTCTGTAACAGGATGAGAAATGTTGACGAGTTACAGAGCTT 3323
QY 288 GluSerLeuGlyMetAspGlnGluThrLeuSerGluValGluAlaIleLeuGluProVal 307
Db 3322 GAAAGTCTGGGATGGATCAAGAAACTCTGTCTGAGTTGAAGCTATTCTCGAGCCTGTA 3263
QY 308 LysAsnLeuThrTyrProSerGlyIleHisGlnAsn 319
Db 3262 AAGATCTGACATGGCCCAAGTGGATCCATCAGAAC 3227

RESULT 9
LOCUS ATCHRV79/c 199536 bp DNA linear PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 79.
ACCESSION AL161583
VERSION AL161583.2 GI:7270260
KEYWORDS Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 65668)
AUTHORS Peters, S.A., van Staveren, M., Dirkse, W., Stiekema, W., Mewes, H.W.,
Lemcke, K., and Mayer, K.F.X.
JOURNAL Unpublished
REFERENCE 2 (bases 64780 to 161254)
```

```
AUTHORS Rose, M., Hempel, S., Entian, K.-D., Mewes, H.W., Lemcke, K. and
Mayer, K.F.X.
JOURNAL Unpublished
REFERENCE 3 (bases 147320 to 199536)
AUTHORS Obermaier, B., Deutschenbaur, S., Piravandi, E., Mewes, H.W., Lemcke, K.
and Mayer, K.F.X.
JOURNAL Unpublished
REFERENCE 4 (bases 1 to 199536)
AUTHORS EU Arabidopsis sequencing project.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
COMMENT Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
this fragment has an overlap with ATCHRV78 at the 5' end and an
overlap with ATCHRV80 at the 3' end.
FEATURES
source Location/Qualifiers
1. 199536
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
7525..8075
/gene="AT4g33130"
complement(join(7525..7590,7667..7852,7950..8075))
/gene="AT4g33130"
complement(join(7525..7590,7667..7852,7950..8075))
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB80030.1"
/db_xref="GI:7270261"
/db_xref="UniProt/TREMBL:Q9SM26"
translation="MIOIKRLQQLQOEINLRALTSAVEHSSSPFMDSPCELPDKEL
LDSALILEITVSKLEQESVSLRYLLRQEKERRLSEILQKKSHSAPSKFTNAQFPN
KSVTRKRESQVDSLDSEALQVE"
complement(7525..7590)
/gene="AT4g33130"
number=1
complement(7591..7666)
/gene="AT4g33130"
number=1
complement(7667..7852)
/gene="AT4g33130"
number=2
complement(7853..7949)
/gene="AT4g33130"
number=2
complement(7950..8075)
/gene="AT4g33130"
number=3
802..12223
/gene="AT4g33140"
join(802..10207,11382..11406,11765..11899,12013..12223)
/gene="AT4g33140"
codon_start=1
/product="hypothetical protein"
/protein_id="CAB80031.1"
/db_xref="GI:7270262"
/db_xref="UniProt/TREMBL:Q9SM25"
translation="MSSFTDPLFLRSLSPRSIAIRFHGDFAHRLYNPIQIYNNNNN
SGSCNGLPKYVGGGKPKGFTLRSFCSHKPKDQKARALQHRFLHDLGSGPSERA
AAGTTLADGRGAYLRFRQNDKIVVAVDIDEMLLSQRRSQNAIKHTLEWIEI
HFPLGKQIHFGNHFHALHGRSPKSEICRSFGAELIDDNPRYAEACANIGMKVLLFD
YENSYPMSKTESVDRLPLVTRVNWVEVEQQLSLAVSKC"
802..10207
/gene="AT4g33140"
exon
intron
exon
intron
exon
gene
CDS
exon
```


Qy 67 MetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAspTyrIleValAlaThr 86
Db 199176 ATGCTTGGTGAAGGACTAAAGGCTTTGCAAGTCCCTAGAGTGACTACATTGGTGGTACT 199117
Qy 87 LysCysGlyArgTyrLysGlyGluGlyPheAspPheSerAlaGluArgValArgLysIle 106
Db 199116 AAGTGTGGTAGATATAAAGAGAGTTTGTGATTTCAAGTGTGAGAGAGTAAAGAGATATT 199057
Qy 107 AspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIleLeuHisCysHisAspIle 126
Db 199056 GACGAGACTTGGAGAGGCTTCAGCTTGATTTATGTTGACATACCTTCATTGGCATGACATT 198997
Qy 127 GluPheGlySerLeuAsp 132
Db 198996 GAGTTCGGGTCTCTTGATCAGGTTTACCACCTTTATTTCTCTTGGGTTTATCATTTCGTT 198937
Qy 133 138
Db 198936 GTTTGTGTCTCAAACTTTTACTAAGTGTTCATTCTGCTCCAGATTGTGAGTGAACA 198877
Qy 139 IleProAlaLeuGlnLysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGly 158
Db 198876 ATTCTGTCTCTCAGAACTGAACAAGAGGGGAGACCCGGTTCATTGGTATCACTGGT 198817
Qy 159 LeuProLeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAspVal 178
Db 198816 CTTCCGTTAGATATTTTCACTATGTTCTTGATGAGTGCCTCCAGGACTGTGATGTG 198757
Qy 179 IleLeuSerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyr 198
Db 198756 ATATTGTCTACTGTTCATTACGGCGTTAATGATTCGAGCTGTGCTGATTTACTACCTTAC 198697
Qy 199 LeuLysSerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThr 218
Db 198696 TTGAAGACAAGGTGGTGTGTAAGTGTCTCTCCATTAGCAATGGGCTCTCTTACA 198637
Qy 219 GluGlnGlyProProGluTrpHisProLysSerProGluLeuLysSer 234
Db 198636 GAACAAGGTCTCTGTAATGGCACCCTGCTCCCTGAGCTCAAGGT-TTGGCATTTTAC 198578
Qy 234 234
Db 198577 AAAAGAATGTGATTTCTGTTAACTTTCTGTAGCATTTGTATTGATACACCGTATAAGAG 198518
Qy 234 234
Db 198517 ACTAGTGTAAACGCCATTCTGTCAGCATTTCTCTGAGCATAGCAGGTAAATTTCTAT 198458
Qy 234 234
Db 198457 AGATGCTATACTTCTGAGACTGACAATCTTTGTTTCTTAACCTGTCGACATCATTTGTGATT 198398
Qy 234 234
Db 198397 AGTCTTCATAGCAAGATTAGAGACATTCAGATAGTTCCCAATGGGTATTAAACATGTTA 198338
Qy 234 234
Db 198337 GTTGTCTTTGATGATGATTTCAGTAGAATAAACAAGATTCAACATGACCGCTGTTATT 198278
Qy 234 234
Db 198277 AGTTGAATTTGATTTTGAAGAGTAGGGAGTCTCTTGAGTTAGTTTGTGTTTGTCTCT 198218
Qy 234 234
Db 198217 GTTTGATATTTTGTATTAGTCAGAGTAAACCAAGATACGTACTCTTTTAAACAAATTTGTC 198158
Qy 235 237
Db 198157 TCAATATGAAGATCATACCGTTGCTGCTTTTATTTATGTTATGATCTGTCGACAGCAA 198098

Qy 238 AlaAlaValAlaHisCysLysSerLysGlyLysLysIleThrLysLeuAlaLeuGlnTyr 257
Db 198097 GCCCAGTGTGCTACTGCAATCAAGGGCAAGAGATCACAAGATTAGCTCTGCAATAC 198038
Qy 258 SerLeuAlaLeuLysGluIleSerSerValLeuValGlyMetSerSerValSer 275
Db 198037 AGTTTAGCAACAAGGAGATTTCGTGCTGTTGGTGGATGAGCTCTGTCTC-ACAGGT 197979
Qy 275 275
Db 197978 ATTCAAGCTTGGTATTCTTCTTGTCAAGTCCCAAAATTATGTGGAGTTGGAGTTAAATAT 197919
Qy 276 287
Db 197918 GAAAGAAATTTGTGTGTTTCTGAAAACAGGTGAAGAAAATGTTGCAGCAGTTTACAGAGTT 197859
Qy 288 GluSerLeuGlyMetAspGlnGluThrLeuSerGluValGluAlaIleLeuGluProVal 307
Db 197858 GAAAGTCTGGGGATGGATCAAGAACTCTGTCTGAGGTTGAGGCTATTCTCGAGGCTGTA 197799
Qy 308 LysAsnLeuThrTrpProSerGlyIleHisGlnAsn 319
Db 197798 AAGAATCTGACATGCCCAAGTGGAAATCCATCAGAAC 197763
RESULT 10
LOCUS AV176585 1215 bp mRNA linear PLN 13-JAN-2003
DEFINITION Actinidia deliciosa galactose dehydrogenase mRNA, complete cds.
ACCESSION AV176585
VERSION AV176585.1 GI:27728730
KEYWORDS
SOURCE Actinidia deliciosa
ORGANISM Actinidia deliciosa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; Ericales; Actinidiaceae; Actinidia.
REFERENCE 1 (bases 1 to 1215)
AUTHORS Beuning, L., Bowen, J., Crowhurst, R., Gleave, A., MacRae, E.,
Newcomb, R., Perera, S., Ross, G., Snowden, K., Walton, E. and Yauk, Y.K.
A Gene Database from Fruit Tree Species
Unpublished
REFERENCE 2 (bases 1 to 1215)
AUTHORS Laing, W.A. and MacRae, E.
TITLE Kiwifruit Leaf Galactose Dehydrogenase
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1215)
AUTHORS Laing, W.A. and MacRae, E.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-2002) Postharvest and Food Science, The
Horticulture and Food Research Institute of New Zealand Ltd, PB
92169, Auckland, New Zealand
FEATURES
source
1..1215
/organism="Actinidia deliciosa"
/mol_type="mRNA"
/cultivar="Hayward"
/db_xref="taxon:3627"
/tissue_type="developing shoot buds from extending water
buds"
32..991
/function="oxidizes L-galactose at C1 to
L-galactono-1,4-lactone"
/codon_start=1
/evidence=experimental
/product="galactose dehydrogenase"
/protein_id="AAO18639.1"
/db_xref="GI:27728731"

ORIGIN

Alignment Scores:
Pred. No.: 1,54e-106 Length: 1215
Score: 1348.00 Matches: 256
Percent Similarity: 90.5% Conservative: 32
Best Local Similarity: 80.50% Mismatches: 30
Query Match: 83.28% Indels: 0
DB: 8 Gaps: 0

US-10-606-300-11 (1-319) x AY176585 (1-1215)

Qy 1 MetThrLysIleGluLeuArgAlaLeuGlyAenThrGlyLeuLysValSerAlaValGly 20
Db 32 ATGACAAACCTAGACCTCGACCCCTCGGAAACACGCGCTTGAAGCTGAGCTCGGTGGGC 91
Qy 21 PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAla 40
Db 92 TTCGGAGGCTCACCTCTGGCAACGCTTTGGCCAGCTCTCGACACGACGCCATCGCC 151
Qy 41 ThrValArgGluAlaPheArgLeuGlyIleAenPheAspThrSerProTyrTyrGly 60
Db 152 TCCGTCGCGAGGCGCTTCGCGCTCGGCATCAACTTCTTCGACACCTCTCCGTATTATGGA 211
Qy 61 GlyThrLeuSerGluLysMetLeuGlyLysGlyLysGlyLysAlaLeuGlnValProArgSer 80
Db 212 GGGACATTTGTCGGAGAAAGTACTGGGGAAGGCACCTAAAGAGCTCTGGGAGTGCCCGCGAAC 271
Qy 81 AspTyrIleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGlu 100
Db 272 GAGTACATTTGTCGACCAAGTGTGGAGGTATGACAGAGGCTTTGATTTTTCAGTCCGAG 331
Qy 101 ArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 120
Db 332 AGAGTGACTAGAGCTTTGATGAGAGCTCGGAGAGGTTGCAGCTCGACTATGTTGATATA 391
Qy 121 LeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePro 140
Db 392 TTGCAATGCCATGATATTGAATTTGGGTCCCTTGATCATGATGTGAATGACAGCATTCCT 451
Qy 141 AlaLeuGlnLysLeuLysGlnGlyLysThrArgPheIleGlyIleThrGlyLeuPro 160
Db 452 GCTCTTCAAAAGCTTAAAGAGACGACGAAGATCCGGTTATTGTTATTACAGGACTCCA 511
Qy 161 LeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeu 180
Db 512 CTGGAGATTTTACCTATGTCTCGATCGGTTCCACGAGGACACAGTCCGACTGATCTG 571
Qy 181 SerTyrCysHisTyrGlyValAsnAspSerThrIleLeuAspLeuProTyrIleLys 200
Db 572 TCATATTGCCACTACAGTATTAAATGATTCAACTCTGGAGGATCTACTGCCCTTACTTAAAG 631
Qy 201 SerLysGlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGln 220
Db 632 AGCAGAGGTGTGGAGTATCATGCTCTCCGCTTGCATGGGCTTCTTACAGAGAGT 691
Qy 221 GlyProGluTyrIleProAlaSerProGluLeuLysSerAlaSerIysAlaVal 240
Db 692 GGCCCTCCAGAGTGGCACCAGCTTCACTGAACCTGAAGCGCGCATGCCAAGCCGCTGCT 751
Qy 241 AlaHisCysLysSerLysGlyLysIleThrLysLeuAlaLeuGlnTyrSerIleAla 260
Db 752 GCTCATTTGTAAGAGAGGGTGAATAATATATCAAGTTAGCTATGATGATGATGATGATGAT 811
Qy 261 AsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGluAsn 280
Db 812 AATAAGATATTCTTCCATAGTGTGGCAGTAATCTGTGTAACAGGTTGAAGAAAT 871
Qy 281 ValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluVal 300
Db 872 GTTGCTGCGCTTAATGAACCTTGCACATTTGGGAAGGATGAGAAACCTGTGTGACAGATT 931
Qy 301 GluAlaIleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGln 318

Db 932 GAAGAGATTCTAAACCCCTGAGAACACGACATGGCTTAGCGGTATACACAA 985

RESULT 11
LOCUS AY264803 1221 bp mRNA linear PLN 30-APR-2003
DEFINITION Malus x domestica L-galactose dehydrogenase mRNA, complete cds.
ACCESSION AY264803
VERSION AY264803.1 GI:30267885
KEYWORDS
SOURCE
ORGANISM Malus x domestica (cultivated apple)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.
REFERENCE 1 (bases 1 to 1221)
AUTHORS Beuning, L., Bowen, J., Crowhurst, R., Gleave, A., Macrae, E.,
Newcomb, R., Perera, S., Ross, G., Snowden, K., Walton, E. and Yauk, Y.K.
A Gene Database from Fruit Tree Species
Unpublished
JOURNAL 2 (bases 1 to 1221)
REFERENCE Laing, W.A. and Macrae, E.
AUTHORS
TITLE Kiwifruit Leaf Galactose Dehydrogenase
JOURNAL Unpublished
JOURNAL 3 (bases 1 to 1221)
REFERENCE Laing, W.A. and Macrae, E.
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (28-MAR-2003) Gene Technologies Sector, The Horticultural
and Food Research Institute of New Zealand, PB 92169, Auckland, New
Zealand
FEATURES
source
1..1221
/organism="Malus x domestica"
/mol_type="mRNA"
/cultivar="Royal Gala"
/db_xref="taxon:3750"
/tissue_type="pre-opened floral bud"
27..1001
/function="oxidizes L-galactose at C1 to
L-galactono-1,4-lactone"
/codon_start=1
/evidence="not experimental"
/product="L-galactose dehydrogenase"
/protein_id="AAP21783.1"
/db_xref="GI:30267886"
/translation="MASQPLPKVBLRGNTGLKSLCVGFGASPLGNVFGVPVSDDEAI
GSVREAFRRGNFFDTSPYQGLTSEKVLGKTLKALGVPRSEYVWPKCKGKIRFIG
SADRVTKSIDESLRLQIDYDILQCHDIEFGSLDQIVNETIPALKKLKEAGKIRFIG
ITGLPLGIFTVLDRI PAGMVDVVL SYCHYGNIDSTLEDLIPYLSKGVGIIISASPLA
MGLITENGPPPEWHFASAEKLSACRAAAVYCKERGNNISKALQYLSLKNKDISSVLVGM
NSINQVBNVAAVELATIGKNEKILAEVEAILKPKVKNQTPWPSGLQQS"

ORIGIN
Alignment Scores:
Pred. No.: 2,67e-104 Length: 1221
Score: 1322.00 Matches: 248
Percent Similarity: 90.28% Conservative: 40
Best Local Similarity: 77.74% Mismatches: 31
Query Match: 81.66% Indels: 0
DB: 8 Gaps: 0

US-10-606-300-11 (1-319) x AY264803 (1-1221)

Qy 1 MetThrLysIleGluLeuArgAlaLeuGlyAenThrGlyLeuLysValSerAlaValGly 20
Db 42 CTGCCAAAGGTGAACCTCGAGAGCTTGGCAACACAGGCGCTCAAGCTGAGTGTGCGC 101
Qy 21 PheGlyAlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAla 40
Db 102 TTCGGGGGCTCCCTCTCGGCAACGCTCTCGGTCCCGTCTCGATGAGCAAGCAATCGGC 161
Qy 41 ThrValArgGluAlaPheArgLeuGlyIleAenPheAspThrSerProTyrTyrGly 60

Db 162 TCTGTCGCGAAGCCTTCGCGGAGCATCAACTCTTTTGACACCTCTCCGTATATATGGA 221
Qy 61 GlyThrLeuSerGluLeuMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSer 80
Db 222 GGGACTTTGTGAGAAAGGCTGTTGGTAAACGCTAAAGCTCTAGGTGTGCCGAGAGC 281
Qy 81 AspTyrIleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGlu 100
Db 282 GAGTACGTTGTGGCAACCAAGTGTGGAGCGCTATGCTGACGGTGTGATTTGATTCAGTGTCTGAT 341
Qy 101 ArgValArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIle 120
Db 342 AGAGTGACTAGAGCATGATGAGAGCTGGAGAGATTGCGAGCTAGATTACGTCGATATATA 401
Qy 121 LeuHisCysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIlePro 140
Db 402 CTGCAATGCCATGACATTGAATTTGGGTCTCTGGACACAGATTGTAAACGAGACATCCCT 461
Qy 141 AlaLeuGlnLysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGlyLeuPro 160
Db 462 GCCTTAAAGAACTGAAAGAGCTGGGAAGATTTCGTTTCATTTGGTATCACAGGACTTCCA 521
Qy 161 LeuAspIlePheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeu 180
Db 522 TTGGGAATTTTTCATATATGTTCTTGTATCGTATACAGCTGGCATGGTGAGCTGTCTCTG 581
Qy 181 SerTyrCysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLys 200
Db 582 TCATATTGCCACTACGGTATTAAATGATTCACATTTGGAGGATCTAATACCATCTCAAG 641
Qy 201 SerLysGlyValGlyLysIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGln 220
Db 642 AGCAAGGTGTGGAAATCATGATGCTCTCCACTTGAATGGGTCTGCTTACTGAGAAT 701
Qy 221 GlyProProGluTyrHisProAlaSerProGluLeuLysSerAlaSerLysAlaVal 240
Db 702 GGTCTCTCAGAAATGGCACCCTGCTCTGCTGAACCTGAAGTCTGCTGGCAGCTCTGCT 761
Qy 241 AlaHisCysLysSerLysGlyLysLysIleThrLysLeuAlaLeuGlnTyrSerLeuAla 260
Db 762 GTCTATTGTAAGAGAGAGGAGGATATATTTCAAGATTAGTCTGCTGCATACAGCTTGTCC 821
Qy 261 AsnLysGluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGln 280
Db 822 AATAAGGATATCTCATGCTGCTGCGCATGAACTCTATTAAATCAGGTTCAGGAAAAAT 881
Qy 281 ValAlaAlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluVal 300
Db 882 GTGCTGCTGCTGCTGCGGAGCTGCGGACCATTTGGCAAGAAATGAGAAATTTCTAGCAGAGTT 941
Qy 301 GluAlaIleLeuGluProValLysAsnLeuThrTyrProSerGlyIleHisGlnAsn 319
Db 942 GAAGCTATTCTGAAACCAAGTGAAGAAATCAGACATGCGCTAGCGGACTCCAGCAGAGC 998
RESULT 12
LOCUS AB160990
DEFINITION Spinacia oleracea GDH mRNA for L-galactose dehydrogenase, complete cds.
ACCESSION AB160990
VERSION AB160990.1 GI:50582466
KEYWORDS Spinacia oleracea (spinach)
SOURCE Spinacia oleracea
ORGANISM Spinacia oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Caryophyllales; Amaranthaceae; Spinacia.
REFERENCE 1
AUTHORS Mieda, T., Yabuta, Y., Rapolu, M., Motoki, T., Ishikawa, T.,
Yoshimura, K. and Shigeoka, S.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1239)
AUTHORS Mieda, T., Yabuta, Y., Madhusudhan, R., Motoki, T., Ishikawa, T.,

Yoshimura, K. and Shigeoka, S.
Direct Submission
Submitted (23-JAN-2004) Yukinori Yabuta, Faculty of Agriculture,
Kinki University, Department of Food and Nutrition; 3327-204
Nakamachi, Nara city, Nara 631-8505, Japan
(E-mail:yabuta@nara.kindai.ac.jp, Tel:81-742-43-7273 (ex.3416),
Fax:81-742-43-2252)
FEATURES
Location/Qualifiers
source 1..1239
/organism="Spinacia oleracea"
/mol_type="mRNA"
/db_xref="taxon:3562"
1..1239
/gene="GDH"
83..1051
/gene="GDH"
/function="oxidises L-galactose at C1, forming
L-galactono-1,4-lactone"
/codon_start=1
/evidence=experimental
/product="L-galactose dehydrogenase"
/protein_id="BAD32687.1"
/db_xref="GI:50582466"
/translation="MNTKLERBELGNTGLNLSCVFGASPLGNVFGDVSEEQSIAT
VIEAFNQINFFDPSFYGATLSKVLGKCLKALGASDEYIVATKCRYAEAGDFSA
ERVTKSIDESLERLQLEYVDLQCHDIEFGSLDQIVNETIPALQIKESGKTRFISIT
GLPEVTTYLDVRYPGTTDVLSYCHYINDSLDMLPYFKGKGVGINASPLSMG
LHTENGPEWHPASPEIKAKAAADYCKKGNKISKLALQYLSNKNIDISTTLVGMNS
VKQVENVGALELETAGKDEKFAEINILKPIKNQSWPSGIQQT"
ORIGIN
Alignment Scores:
Pred. No.: 4.8e-99 Length: 1239
Score: 1261.00 Matches: 237
Percent Similarity: 87.03% Conservative: 38
Best Local Similarity: 75.00% Mismatches: 41
Query Match: 77.89% Indels: 0
DB: 8 Gaps: 0
US-10-606-300-11 (1-319) x AB160990 (1-1239)
Qy 3 LysIleGluLeuArgAlaLeuGlyAsnThrGlyLeuLysValSerAlaValGlyPheGly 22
Db 98 AAATTAGAGCCCGAGAAATTCGGAAACACTCGCCTTAACCTCAGTGTGTGGGTTCGCG 157
Qy 23 AlaSerProLeuGlySerValPheGlyProValAlaGluAspAlaValAlaThrVal 42
Db 158 GCCTCTCTCTTGGCAAGCTTTTCGGTATGTTTCTGAGGAACAATCCATTGCCACTGTT 217
Qy 43 ArgGluAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyrTyrGlyGlyThr 62
Db 218 ATTGAAGCTTTTAATCAGGGTATCAACTTCTTTGACACTTCTCCGTACTATGGTCCACA 277
Qy 63 LeuSerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAspTyr 82
Db 278 TTGTGCGAAGAGTCTTGGGAAGTGTTTGAAAGCACTTGTGTGATCATAGATGAATAC 337
Qy 83 IleValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGluArgVal 102
Db 338 ATTGTTGCAACGAATATGCGCGCTTATGCCGAGGCTTTGATTTTAGTGTGAAGAGTG 397
Qy 103 ArgLysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIleLeuHis 122
Db 398 ACTAAGACATTTGATGAAGCTAGAAAGCTAGAAAGTTGCAACTTGATGTATGATATACAA 457
Qy 123 CysHisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIleProAlaLeu 142
Db 458 TGTCAATGATATGAATTTGGTCTCTTGTATGATTTGCAATGTCAATGAGACTATTCCGCCCTG 517
Qy 143 GlnLysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGlyLeuProLeuAsp 162
Db 518 CAGAAATTAAGGAATCATCGGAAGACTAGATTCATTTGGTATTAATCTGGACTTCCATTTGNA 577

QY 163 llePheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeuSerTyr 182
 Db 578 GGTGTACAGTATGTTCTTGACCGTGTCCCGCTGGGCAATGATGTGTGTTCTGTCATAT 637
 QY 183 CysHisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuIlysSerLys 202
 Db 638 TGCCACTATTGTATCAATGACTCACTCTGGAGACATGCTACCGTATTTTCAGAGCAAA 697
 QY 203 GlyValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGlnGlyPro 222
 Db 698 GCGCTGTGTGTGATCAATGCTTCTCCACTTTCAATGGGCTTCATACAGAAATGGCCCC 757
 QY 223 ProGluThrHisProAlaSerProGluLeuIlysSerAlaSerLysAlaAlaValAlaHis 242
 Db 758 CCAGAGTGGCATCCAGCTTCACCTGAGATTAAAGCTGCGATGCAAGCTGCTGCAGATTAC 817
 QY 243 CysLysSerLysGlyLysLysIleThrLysLeuAlaLeuGlnTyrSerLeuAlaAsnLys 262
 Db 818 TGTAAGAGAATGGAGAGATATCTCAAAATTAGCTTCGCAGTACAGCTGTCAACAAA 877
 QY 263 GluIleSerSerValLeuValGlyMetSerSerValSerGlnValGluGluAsnValAla 282
 Db 878 GACATTTCCACTACTCTTTGTCATGAACCTCTGTTAAGCAGGTGGAGAAATGTTGGT 937
 QY 283 AlaValThrGluLeuGluSerLeuGlyMetAspGlnGluThrLeuSerGluValGluAla 302
 Db 938 GCGCGTTTAGAACTTGAACCTCGACGAAGGATGAGAAACATTTGCGAGAAATTCAGAAC 997
 QY 303 lleLeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGln 318
 Db 998 ATCTTAAGCCTTAAAGATCAAGTTGGCCCTAGTGTATACACAA 1045

RESULT 13
 AK060431
 LOCUS Oryza sativa (japonica cultivar-group) cDNA clone:001-011-B02, full
 DEFINITION insert sequence.
 ACCESSION AK060431
 VERSION AK060431.1 GI:32970449
 KEYWORDS FLI CDNA; oligo-capping.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.

REFERENCE 1
 AUTHORS The Rice Full-Length cDNA Consortium, National Institute of
 Agrobiological Sciences Rice Full-Length cDNA Project Team:
 Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
 Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
 Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
 Ohtsuka,K., Shishiki,T., Foundation of Advancement of International
 Science Genome Sequencing & Analysis Group: Ootomo,Y., Murakami,K.,
 Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
 Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
 Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J.,
 Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
 Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:
 Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S.,
 Hara,A., Hashizume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,
 Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,
 Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
 Yoshino,M. and Hayashizaki,Y.
 Collection, mapping, and annotation of over 28,000 cDNA clones from
 japonica rice
 Science 301 (5631), 376-379 (2003)

TITLE japonica rice
 JOURNAL Science 301 (5631), 376-379 (2003)
 MEDLINE 22752273
 PUBMED 12869764
 REFERENCE 2 (bases 1 to 1234)
 AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,
 Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W.,
 Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T.,
 Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K.,

Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,
 Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M.,
 Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,
 Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
 Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,
 Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,
 Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M.,
 Nishiki,T., Nariikawa,R., Niikura,J., Nishi,K., Nomura,K.,
 Numasaki,R., Ohneda,E., Ohno,M., Ohtsuka,K., Oka,M., Ooka,H.,
 Osato,N., Ota,Y., Ootomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K.,
 Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K.,
 Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S.,
 Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y.,
 Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A.,
 Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W.,
 Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and
 Yoshimura,A.
 Direct Submission
 Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of
 Agrobiological Sciences, Department of Molecular Genetics, Head of
 Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
 305-8602, Japan (E-mail: skikuchi@niae.affrc.go.jp,
 Tel:81-29-838-7007, Fax:81-29-838-7007)

COMMENT This clone is one of the 28K full-length cDNA clones from japonica
 rice.

URL : http://cdna01.dna.affrc.go.jp/cDNA/
 NIAS Rice Full-length cDNA Project Team: Kikuchi,S., Satoh,K.,
 Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,
 Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T.,
 Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuka,K., Shishiki,T. and
 Yamamoto,M.
 FAIS Genome Sequencing & Analysis Group: Ootomo,Y., Iida,Y.,
 Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M.,
 Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J.,
 Mizuno,K., Narikawa,R., Niikura,J., Oka,M., Ryu,R., Sugano,S.,
 Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S.,
 Yoshimura,A., Matsubara,K. and Murakami,K.
 Genome Exploration Research Group in Riken Genomic Sciences Center
 and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K.,
 Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T.,
 Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K.,
 Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y.,
 Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J.,
 Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
 Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
 Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N.,
 Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H.,
 Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
 Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F.,
 Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K.,
 Yasunishi,A. and Hayashizaki,Y.
 Location/Qualifiers
 1..1234
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultiivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="001-011-B02"

FEATURES

source

ORIGIN

Alignment Scores:
 Pred. No.: 4,328-95 Length: 1234
 Score: 1215.00 Matches: 235
 Percent Similarity: 83.49% Conservative: 28
 Best Local Similarity: 74.60% Mismatches: 52
 Query Match: 75.05% Indels: 0
 DB: 8 Gaps: 0

US-10-606-300-11 (1-319) x AK060431 (1-1234)

QY 4 lleGluLeuArgAlaLeuGlyAsnThrGlyLeuIysValSerAlaValGlyPheGlyAla 23

Db 72 ATGGAGCTCCGCGAGCTCGGCGCCACCGGCTCCGCGTCCGCGCTCGGCTTCGGCGGC 131

Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N., Ota,Y., Saitoh,H., Sakai,K., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Takami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayaeshizaki,Y.

FEATURES
source location/Qualifiers
1..1305
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="J01309LL23"

ORIGIN
Alignment Scores:
Pred. No.: 4,65e-95 Length: 1305
Score: 1215.00 Matches: 235
Percent Similarity: 83.49% Conservative: 28
Best Local Similarity: 74.60% Mismatches: 52
Query Match: 75.05% Indels: 0
DB: 8 Gaps: 0
US-10-606-300-11 (1-319) x AK067039 (1-1305)
QY 4 IIEGLUeUatqAlaLeuGlyAsnThrClyLeuLysValSerAlaValGlyPheGlyAla 23
Db 57 ATGGAGCTCCGGAGCTCGGCCACAGGCGCTCCGCGTCAGCCCCGTCGCGCTTCGGCGCC 116
QY 24 SerProLeuGlySerValPheGlyProValAlaGluAspAlaValAlaThrValArg 43
Db 117 TCCCTCTCGGCCATGTCTTCGGGAGCTCCCCCGCGAGCTCGCCCGCGCGCGTCGCG 176
QY 44 GluAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyrTyrGlyGlyThrLeu 63
Db 177 CGCGGCCCTCGACCTCGGATCAACTTCTTCGACACCTCCCGCTACTACGGCGGCACGGT 236
QY 64 SerGluLysMetLeuGlyLysGlyLeuLysAlaLeuGlnValProArgSerAspTyrIle 83
Db 237 TCGAGTCGGTGCTCGGGAGCTGCGCTCGCGCCGGCGGTGCGCGGAGCCGTTCTGTC 296
QY 84 ValAlaThrLysCysGlyArgTyrLysGluGlyPheAspPheSerAlaGluArgValArg 103
Db 297 GTGCCACCAAGTGGCGGCGGTACAGGAAGGGTTTCGACTTCAGCGCGCCCGCGTCACC 356
QY 104 LysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIleLeuHisCys 123
Db 357 CGAGCGTCGAGAGAGCTTCGACGCTTCGGCTTCGACTACGTCGACATCTCTCACTGC 416
QY 124 HisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIleProAlaLeuGln 143
Db 417 CAGCATCGAGTTCACCGACCTCGACAGATTGTGAATGAGACGATTCCGGTGCTCCAG 476
QY 144 LysLeuLysGlnGluGlyThrArgPheIleGlyIleThrGlyLeuProLeuAspIle 163
Db 477 AAGATTCAGGAGCGGGAAGCGCGGTTCTATCGGGATAACCGGGCTGCCATTGACGATC 536
QY 164 PheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeuSerTyrCys 183
Db 537 TACACTTATGTCTCGACAGGTGCGGCAGGTTCTGTGGATGTGATTCTGTCTTACTGC 596
QY 184 HisTyrGlyValAsnAspSerThrLeuLeuAspLeuLeuProTyrLeuLysSerLysGly 203
Db 597 CACTATGGGATCAATGATACCGCGCTTGTGGAATTTGTTCTCCCTACCTGAAGAGCAAGGT 656
QY 204 ValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGlnGlyProPro 223
Db 657 GTTGGCGTATCAGTGCTTCACTCTTCGAATGGGGCTTCTTACGATAACGGGCCACCC 716
QY 224 GluTrpHisProAlaSerProGluLeuLysSerAlaSerLysAlaValAlaHisCys 243
Db 717 GAATGGCACCTTCACCGAAGAACTAAAGTTGGCATCGGGCAGGACGATCACTGT 776

Qy	244	IysSerIysGlyIysIysIleThrIysLeuAlaLeuGlnTyrSerLeuAlaAenIysGlu	263
Db	777	AAAGAAGAAAGGGGAAAAACATTACAAAGAGCTAGCTATGCAGCTACAGCTTGTATGAACAATGAG	836
Qy	264	IleSerSerValLeuValGlyMetSerSerValSerGlnValGluGluAenValAlaAa	283
Db	837	ATTTCACGGTCTCTGTTGGATGAATCACTCTCCAGAACAGGTAGAGAGATGTGGCTGCT	896
Qy	284	ValThrGluLeuGluSerLeuGlyMetAspGlnGluThrIleuSerGluValGluAlaIle	303
Db	897	GCAATTGAATTGCTACTCTCCGGTATTGATTAAGAAACTTCTGCACGAAGTTGAAGCAATT	956
Qy	304	LeuGluProValIysAsnLeuThrTrpProSerGlyIleHisGln	318
Db	957	CTTGAGCCTGTCTCAAGAACATGACCTCGTCCAGTGGCATTGAGCAA	1001
RESULT 15			
AKI02223			
LOCUS	AKI02223	1275 bp	mRNA linear PLN 24-JUL-2003
DEFINITION	Oryza sativa (japonica cultivar-group) cDNA clone:J033087M07, full insert sequence.		
ACCESSION	AKI02223		
VERSION	AKI02223.1 GI:32987432		
KEYWORDS	FLI CDNA; CAP trapper.		
SOURCE	Oryza sativa (japonica cultivar-group)		
ORGANISM	Oryza sativa (japonica cultivar-group)		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.		
REFERENCE	1		
AUTHORS	The Rice Full-length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-length cDNA Project Team; Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group; Otomo,Y., Murakami,K., Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y., Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M., NariKawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Niikura,J., Ikeda,R., Ishibiki,J., Kawamura,M., Yoshimura,A., Miura,J., Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., KIKEN: Kawai,J., Carninci,P., Adachi,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hahidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y., Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Yoshino,M. and Hayashizaki,Y.		
TITLE	Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice		
JOURNAL	Science 301 (5631), 376-379 (2003)		
MEDLINE	22752273		
PUBMED	12869764		
REFERENCE	2 (bases 1 to 1275)		
AUTHORS	Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K., Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K., Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I., Kikuchi,S., Kishikawa,H., Kawagashira,N., Kawai,J., Kawamura,M., Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M., Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M., Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A., Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M., Namiki,T., NariKawa,R., Niikura,J., Nishi,K., Nomura,K., Nomura,K., Nomaeaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H., Osato,N., Ota,Y., Otomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K., Sakaume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Taya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W.,		

Yanada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.
 Direct Submission
 Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: skikuchienias.affrc.go.jp, Tel: 81-29-838-7007, Fax: 81-29-838-7007)
 This clone is one of the 28K full-length cDNA clones from japonica rice.

COMMENT

URL : http://cdna01.dna.affrc.go.jp/cDNA/
 NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and Yamamoto, M.
 FAIS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Norikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.
 Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Kato, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.
 Location/Qualifiers
 1. 1275
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cuiovar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="J033087M07"

ORIGIN

Alignment Scores:
 Pred. No.: 9,96e-95 Length: 1275
 Score: 1211.00 Matches: 234
 Percent Similarity: 83.1% Conservative: 28
 Best Local Similarity: 74.29% Mismatches: 53
 Query Match: 74.80% Indels: 0
 Ds: 8 Gaps: 0

US-10-606-300-11 (1-319) x AK102223 (1-1275)

Qy	4	IleGluLeuArgAlaLeuGlyAsnThrGlyLeuTysValSerAlaValGlyPheGlyAla	23
Ds	34	ATGAGCTCCGGAGCTCGGCCACCGGCTCCGGCTCAGCCCGTTCGGCGCC	93
Qy	24	SerProLeuGlySerValPheGlyProValAlaGluAspAlaValAlaThrValArg	43
Ds	94	TCCCTCTCGGCATGTCTCGGCACGTCCTCCCGCAGCTCGCCCGCGCGCGTCCGC	153
Qy	44	GluAlaPheArgLeuGlyIleAsnPhePheAspThrSerProTyrTyrGlyThrLeu	63
Ds	154	CGCGCTCGACCTCGGCATCACTCTTCGACATCTCCCGTACTACGGCGCACGGTG	213
Qy	64	SerGluLysMetLeuGlyLysGlyLeuLysAlaLeuValProArgSerAspTyrIle	83
Ds	214	TCGGAGTCGGTCTCGGGAGCTGCCTCCGGCGCGGGCGTCCCGGACCGGTTCTGTG	273
Qy	84	ValAlaThrLysCysGlyArgTyrTyrGlyGluPheAspPheSerAlaGluArgValArg	103

Ds	274	GTCCGCCACCAAGTCGGCGCTACAGGAGAGGGTTGACTTCAGCGCGCGCGGTCCACC	333
Qy	104	LysSerIleAspGluSerLeuGluArgLeuGlnLeuAspTyrValAspIleLeuHisCys	123
Ds	334	CGCAGCGTCGACGAGAGCGCTCGCAGCGCTCGACCTCGACATCGTTCACCTCCACATGC	393
Qy	124	HisAspIleGluPheGlySerLeuAspGlnIleValSerGluThrIleProAlaLeuGln	143
Ds	394	CACGACATCGAGTTTCACCGACCTCGACGATTTGTGAATGAGACGATTCGGTGTCTCCAG	453
Qy	144	LysLeuLysGlnGluGlyLysThrArgPheIleGlyIleThrGlyLeuProLeuAspIle	163
Ds	454	AAGATCAAGGACGCGGAGCGGTTTCATCGGATTAACCGGCTGCCATTGAGCATC	513
Qy	164	PheThrTyrValLeuAspArgValProProGlyThrValAspValIleLeuSerTyrCys	183
Ds	514	TACACTTATGTGCTCGACCGAGTGTGTGGATTTGCTTGTGATTTCTCTTACTGTC	573
Qy	184	HisTyrGlyValAsnAspSerThrLeuLeuAspLeuProTyrLeuLysSerLysGly	203
Ds	574	CACATATGGGATCAATGATACCGCGCTTGTGGATTTGCTTACCTTACAGTAAACGCGCACCC	633
Qy	204	ValGlyValIleSerAlaSerProLeuAlaMetGlyLeuLeuThrGluGlnGlyProPro	223
Ds	634	GTTCGGTGATCAGTGTTCACCTTCGCAATGGGCTTCTTACAGTAAACGCGCACCC	693
Qy	224	GluTrpHisProAlaSerProGluLeuLysSerAlaSerLysAlaAlaValAlaHisCys	243
Ds	694	GAATGGCACCTCGCACCGAAGAACTAAAGTTGGCATCGCGGACGACAGATCACTGT	753
Qy	244	LysSerLysGlyLysLysIleThrLysLeuAlaLeuGlnTyrSerLeuAlaAsnLysGlu	263
Ds	754	AAGAAGAAGGGGAAAAACATTACAAAGCTAGCTATGCAGTACAGCTTGATGAACAATGAG	813
Qy	264	IleSerSerValLeuValGlyMetSerSerValSerGlnValGluGluAsnValAlaAla	283
Ds	814	ATTTCCACGGTCTTGTGGATGAATGAATCTCCAGAACAGGTAGAGGAGATGTGGCTGCT	873
Qy	284	ValThrGluLeuGluSerLeuGlyMetAspGlnGlnThrLeuSerGluValGluAlaIle	303
Ds	874	GCAATTGAATTGTCTACTTCCGGTATTGATAAAGAACTTCTGCACGAAAGTTGAAGCAATT	933
Qy	304	LeuGluProValLysAsnLeuThrTrpProSerGlyIleHisGln	318
Ds	934	CTTGAGCCTGTCAAGAACAATGACCTGCTGCTCCAGTCCGATTTGAGCAA	978

Search completed: November 13, 2005, 10:15:53
 Job time : 5076 secs

THIS PAGE BLANK (USPTO)